(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
2 September 2004 (02.09.2004)

PCT

(10) International Publication Number
WO 2004/074429 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number:
PCT/DK2004/000117

(22) International Filing Date: 23 February 2004 (23.02.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

PA 2003 00268	21 February 2003 (21.02.2003)	DK
PA 2003 00269	21 February 2003 (21.02.2003)	DK
60/448,480	21 February 2003 (21.02.2003)	US
60/448,460	21 February 2003 (21.02.2003)	US
PA 2003 01356	18 September 2003 (18.09.2003)	DK
60/504,748	22 September 2003 (22.09.2003)	US

(71) Applicant *(for all designated States except US)*: NUEVOLUTION A/S [DK/DK]; Rønnegade 8,5. sal, DK-2100 København Ø (DK).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **FRESKGÅRD, Per-Ola** [SE/SE]; Örtugsgatan 50, S-603 79 Norrköping (SE). **GOULIAEV, Alex, Hauhr** [DK/DK]; Brødsted 223, DK-3670 Veksø Sjælland (DK). **THISTED, Thomas** [DK/DK]; Fjordskrænten 14, DK-3600 Frederikssund (DK). **OLSEN, Eva, Kampmann** [DK/DK]; Vingetoften 17, DK-2730 Herlev (DK).

(74) Agent: **HØIBERG A/S**; Store Kongensgade 59 A, DK-1264 Copenhagen K. (DK).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declaration under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW, ARIPO patent (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)*

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: METHOD FOR PRODUCING SECOND-GENERATION LIBRARY

(57) Abstract: The present invention relates to a method for generating a second-generation library. In a first step, a library of encoded molecules associated with an identifier nucleic acid comprising codons identifying chemical entities that have participated in the formation of the encoded molecule is provided. In a second step, the library is partitioned and encoded molecules having a certain property are selected. Codons of identifiers of selected encoded molecules are subsequently identified, and a second-generation library is prepared using at least some of the chemical entities coded for by the identified codons. The new focussed library may be used for another partition step to select encoded molecules with a certain property.

WO 2004/074429 A2

Title

METHOD FOR PRODUCING SECOND-GENERATION LIBRARY

Various patent and non-patent references cited in the present application are hereby incorporated by reference in their entirety.

Technical Field of the Invention

The invention relates to a method for producing a second-generation compound library with an improved desired property profile. In nature and artificial methods based on the natural system, the parent genotype is carried on to the off-spring and results in a phenotype in which the exact type and sequence of amino acids is retained, unless a mutation and/or recombination has occurred. The present method only retains the identity of chemical entities, e.g. amino acids, while the sequence wholly or partly is scrambled. The result is a focused second-generation library with lower diversity.

Background of the Invention

The biological evolution is based on the survival of specific genotypes that encode phenotypes with the most suitable functionalities in a certain environment. In all living species DNA programs the genotype. DNA serves two important functions in the natural selection process. One function is obviously to encode for the type of nucleotides used and the other function is to encode for the specific order of nucleotide sequences in a nucleic acid sequence. The strategy used in nature, i.e. encoding for the exact type as well as the precise sequence of nucleotides, ensures an extremely similarity between the progeny and its parents. Thus, conserving almost the exact sequence and type of the nucleotides is absolutely essential in order to create off spring with a high functionality. The changes in the genotype from one generation to another, which allow for evolution, are determined by the random mutation rate and recombination between the two parent's genotypes.

The natural selection cannot afford too many changes in the DNA from one generation to the next in order to secure survival of the species. Therefore, nature has evolved sophisticated means to proofread the copying of the DNA from the parents to its progeny and secured that the characteristics of phenotype from one generation to the next is carried only by the DNA.

Within the art of selecting ligands from a library of encoded polypeptides associated with a corresponding identifier nucleic acid sequence, the method of nature is used. Thus, when more than a single library generation is needed, the identifier nucleic acid sequences (genotype) carries the information from one generation to the next.

WO 93/03172 A1 discloses a method for identifying a polypeptide ligand having a desired property in a polypeptide library. In a first step, a translatable mRNA mixture is provided, which is mixed with a mixture of ribosome complexes to form a translation product attached to the mRNA strand responsible for the formation thereof. In a second step the ribosome complexes binding to a target are partitioned from and remainder of the library. In a third step, an amplification of mRNA strands of the partitioned ribosome complexes, which has bound to the target follows. The amplified mRNA strands are used for the production of a second generation library, which is subjected to a renewed contact with the target. The method is repeated a sufficient number of times until the size of the library has narrowed to a small pool of high affinity binders.

In WO98/31700 A1 a method for selecting a DNA molecule, which encodes for a desired protein, is disclosed. The method implies the initial presence of a pool of candidates RNA molecules, which subsequently is translated into a corresponding pool of RNA-protein fusions. Subsequently the mRNA-protein fusion products are subjected to a selection process, i.e. the fusion products are presented for a target molecule, and a new pool of complexes capable of binding to the target are partitioned. From the new pool of complexes, the mRNAs are recovered and amplified for use in a subsequent round of library generation. Xu, L. et al *Chemistry & Biology, Vol. 9, 933-942, August 2002* discloses a practical embodiment in which a library of more than 10^{12} unique mRNA-protein fusion products through ten rounds of library generation and selection are used to identify a high affinity binding protein.

The preparation of libraries of synthetic molecules associated with a corresponding identifier nucleic acid sequence, and the selection of synthetic molecules from such libraries, have been the subject of various patent applications. When two or more generations of libraries are needed, the identifier nucleic acid sequence is used as carrier between an initial library and the next generation library.

Thus, in WO 00/23458A1 libraries of complexes comprising non-natural molecules attached to corresponding nucleic acid sequences are suggested. After a selection of the library has been conducted, the nucleic acid sequences of successful complexes are amplified by PCR and a new library is prepared from these nucleic acid sequences. The same method of carrying information from an initial library to the next library is applied in WO 02/074929A2 and WO 02/103008A2.

The present invention provides a new method for evolving encoded molecules. The method is based on the identification of chemical entities used in the synthesis of reaction products of successful complexes and the application, at least in part, of these chemical entities in the preparation of the next generation library. The utilization of preferable chemical entities and the exclusion of certain undesired chemical entities in the next library generation generally imply that the next generation library has a smaller size compared to the size of the initial library, thereby, at the same time, retaining the desirable encoded molecules in the library.

Summary of the Invention

The present invention concerns a method for producing a composition of molecules with an improved desired property, said method comprising the steps of: providing an initial library comprising a plurality of different encoded molecules associated with a corresponding identifier nucleic acid sequence, wherein each encoded molecule comprises a reaction product of multiple chemical entities and the identifier nucleic acid sequence comprises codons identifying said chemical entities; subjecting the initial library to a condition partitioning members having encoded molecules displaying a predetermined property from the remainder of the initial library; identifying codons of the identifier nucleic acid sequences of the partitioned members of the initial library; and preparing a second-generation library of encoded molecules using the chemical entities coded for by the codons of the partitioned members of the initial library or a part thereof.

The present invention relates to a novel approach to perform evolution of molecules with a desired property, said approach being different from the approach of nature and the prior art. The invention is based on the selecting of chemical entities, the counterpart of amino acids in Nature, instead of the precise sequence of chemical

entities. This new approach is powerful in *ex vivo* conditions when high functionality of the off spring is not vital for success and when the number of chemical entities relative to the number of reactants used in each encoded molecule is high.

The method disclosed herein will be increasingly effective as the library size increases. This is due to the fact that more chemical entities is used when a library size is increased, when the number of reactions for the formation of the encoded is fixed and the fact that different chemical entities tend to be involved in encoded molecules having the desired property. The chemical entities, which are part of the final selected molecules, will be enriched in each round of selection. Finally, when the diversity has been extensively reduced, the enriched molecules are decoded from the identifier nucleic acid sequence comprising the codons of the chemical entities that have participated in the formation of the encoded molecule.

The strategy of performing enrichment of chemical entities instead of specific combinations of chemical entities more efficiently search the chemical space for all combinations of chemical entities that are eager to show a certain property, such as a binding ability towards a target. Thus, chemical entities having a certain impact on the formation of encoded molecules is allowed in a new library to recombine in each new library generation. In a certain aspect of the invention, the recombination is random, i.e. once a chemical entity has qualified as being of interest it is allowed in every position of the reaction sequence. In another aspect of the invention, the recombination is semi-random, i.e. once a chemical entity is qualified as being of interest it is used in a certain position in the reaction sequence of the encoded molecule. In still a further aspect of the invention, the amount of the chemical entity used in a subsequent library generation is dependent on the frequency and the amount of the partitioned library members.

The present invention may be of special interest when a group of chemical entities are selected from a larger pool of chemical entities in the formation of a first library. Selecting chemical entities resulting in encoded molecules having a certain property in a first library and spiking with remaining chemical entities of the pool allows for the formation of a second-generation library not necessarily of a smaller size but enriched in encoded molecules having a certain property.

The second-generation library may be formed of a reaction product of the chemical entities without attaching the reaction product to a nucleic acid. In an embodiment of such second-generation library the individual reaction products are formed in discrete reaction compartments in accordance with traditional combi-chem technology.

In a certain aspect of the invention, the second-generation library is prepared as the first generation library, i.e. the second-generation library comprises a plurality of different encoded molecules associated with a corresponding identifier nucleic acid sequence, wherein each encoded molecule comprises a reaction product of multiple chemical entities and the identifier nucleic acid sequence comprises codons identifying said chemical entities.

In a preferred aspect of the invention, it comprises subjecting the second-generation library to a condition partitioning members having encoded molecules displaying a predetermined property from the remainder of the second-generation library. The second-generation library may be partitioned as to the same property or a different property. Notably, the second-generation library can be screening against the same target or a different target.

After the partitioning of the second-generation library, the invention comprises the step of deducing the identity of the encoded molecule(s) using the identifier nucleic acid sequence, when present. Optionally, a third or further generation library may be formed and screened before the final deducing step is performed. In a certain embodiment, the decoding includes that the codons of the identifier nucleic acid sequence is decoded to establish the synthesis history of the encoded molecules. The synthesis history includes the identity of the chemical entities used and the point in time they enter the sequence of reactions resulting in the encoded molecule.

The encoded molecule is preferably a reaction product in which multiple chemical entity precursors have participated. The encoded molecule may have any chemical structure. Generally, the multiple chemical entities are precursors for a structural unit appearing in the encoded molecule. However, the chemical entities may also perform a chemical reaction with the nascent encoded molecule, which result in an altering or removal of chemical groups. In certain aspects of the invention, the encoded molecule is a scaffolded molecule, i.e. various chemical entities have reacted with a chemical core structure like steroid, benzodiazepine, retinol,

camphor, ephedrine, penicillin, cannabinol, coumarin, oxazol, etc. In certain other aspects of the invention the encoded molecule is fully or partly a polymer. The polymer may be of a type which occurs naturally or may be a non-naturally occurring polymer. Nature only has the possibility of preparing α-polypeptides using the recognition of a codon of an mRNA strand by the anticodon of a charged tRNA. In some aspects of the invention, the encoded molecule is not a α-peptide. Notably, in some aspects of the invention, the chemical entities are reacted without enzymatic interaction to produce the encoded molecule.

The encoded molecule can be associated with the nucleic acid sequence identifier in any appropriate way. In a certain aspect of the invention, the encoded molecule associated with the corresponding identifier nucleic acid sequence is a bifunctional complex. The bifunctional complex may be formed by covalent or non-covalent attachment of the encoded molecule to the identifier nucleic acid sequence. In another aspect of the invention, an identifier nucleic acid sequence is physically a distinct entity separated from the encoded molecule, wherein the identifier identifies the spatial position of an encoded molecule, e.g. in the same compartment in which an encoded molecule is formed a corresponding identifier oligonucleotide is generated.

The conditions partitioning complexes of interests from the remainder of the library may be chosen from a variety of possibilities. In one aspect the condition relates to physical parameters, so that complexes displaying a physical stability under e.g. certain temperature conditions, certain acidic conditions, certain radiation conditions etc. are selected from the library. In other aspects of the invention the condition for partitioning the desired complexes includes subjecting the initial library to a molecular target and partitioning complexes binding to this target. The molecular target may be any compound of interest. Exemplary targets are proteins, carbohydrates, polysaccharides, hormones, receptors, antibodies, viruses, antigens, cells, tissues etc. In certain aspects the target is immobilized on a solid support, such as column material and contacted with the candidate complexes in a fluid media followed by a partitioning of the complexes capable of binding to the target under the contacting conditions used. Typically the binding complexes are eluted from the column using increased stringency conditions.

The complexes as such or only the identifier part is harvested after the partitioning step. Usually the identifier nucleic acid sequences are amplified prior to the identification step. The amplification is suitably performed applying polymerise chain reaction (PCR). The amplified identifiers may be explicitly or implicitly identified. When the codons are identified explicitly, the sequence and identity of nucleotides in the codon is made known to the experimenter, whereas, when the codons of the identifiers are implicitly identified, the experimenter is not presented for the information.

Any suitable method for identifying codons may be used. In a certain aspect of the invention, traditional sequencing, e.g. by using a modification of the Sangers method or pyrosequencing methods, identifies the codons. In another aspect of the invention, the codons of the identifier nucleic acid sequences of the partitioned members of the initial library are identified by contacting said identifier nucleic acid sequences with a pool of nucleic acid fragments under conditions allowing for hybridisation.

The pool of nucleic acid fragments may be immobilized or in solution. In a certain aspect of the invention, the pool of nucleic acid fragments comprises a plurality of single stranded nucleic acid probes immobilized in discrete areas of a solid support, wherein the nucleic acid probes are capable of hybridising to a codon of the identifier nucleic acid sequence comprising codons. The nucleic cid probes may be positioned on a microarray, such that the identity of the codons is revealed by observing the discrete areas of the support in which a hybridisation event has occurred.

The nucleic acid probe can be directly hybridised to the identifier or the nucleic acid probe of the array is hybridised to an identifier nucleic acid sequence through an adapter oligonucleotide having a sequence complementing the probe as well as one or more codons of the identifier nucleic acid sequence. The probe may identify a single codon of an identifier or a probe of the array is capable of hybridising to two codons of the identifier nucleic acid sequence or a sequence complementary to said sequence. The ability to hybridise two or more codons makes it possible to study the influences of neighbouring chemical entities on each other. In a certain aspect, a nucleic acid probe of the array is capable of hybridising to all codons of an identifier nucleic acid sequence. This latter option will fully decode the identity of the encoded

molecule. Usually however, a fully decoding is only possible for a relative small library size, as it presupposes a nucleic acid probe for each member of the library.

When single codons are detected, useful information about a certain codon may be gathered by detecting the codon together with a framing sequence identifying the position in the reaction history of the chemical entity corresponding to said codon.

As an example, if a library of complexes is prepared from 100 chemical entities and the three reactions, i.e. each identifier comprises 4 codons, the library size is 10^8. For most practical uses 10^8 is in the excess of what is possible to detect on an array, especially if multiple determinations for each identifier are considered necessary to obtain a high accuracy. However, an array of just 100 probes complementary to the 100 codons will reveal important information prior to or subsequent to a selection. In the event a framing sequence is detected together with the codon an array of 400 probes is needed.

A suitable method for identifying an hybridisation event is to use a label. Therefore, in a preferred embodiment, the existence of a hybridisation event is measured through labelling of the identifier nucleic acid sequence, or an amplification product thereof. When the label emits light, the hybridisation event is measured by the emission of light in a scanner. To reveal the relative abundance of each chemical entity in the library of encoded molecules, the relative intensity of light in each discrete spot is measured.

The measurement of a hybridisation event may be conducted by various methods known in the art. In the event the label emits lights, the presence or absence of a hybridisation event may be measured in a scanner, e.g. a confocal scanner. The scanner may be connected with computer software, which is able to quantify the amount of lights measured. The amount of light measured correlates with the amount of identifier annealed to the probes. Thus, it is possible to measure not only the presence or absence of one or more codons of an identifier ; it is also possible to measure the relative amount of the codons in one or more identifiers.

After the complexes have been partitioned and the specific codons have been identified on the microarray, the information can be used to design optimized libraries

including chemical entities based on both the selection data and the chemical structure. The microarray analysis will first of all detect which chemical entities pass the partitioning step. Secondly, the relative intensity on the microarray will reflect the relative binding affinity of the chemical entities. Finally, the structures of the chemical entities are directly identified due to the position of the probes on the array. For instance, chemical entities that are strongly selected in a partitioning process but possess some unfavourable chemical structure can be excluded in the next generation of library. Similarly, chemical entities that are weekly selected in a partitioning process but possess some favourable chemical structure can be included in the next generation of library. Thus, the next generation library design can be based both on a rational choice of chemical entities with lead-like structures and the selection pressure detected on the microarray.

Another method of identifying codons includes that nucleic acid fragments are primer oligonucleotides, and the identification involves subjecting the hybridisation complex between the primer oligonucleotides and the identifier nucleic acid sequences to a condition allowing for an extension reaction to occur when the primer is sufficient complementary to a part of the identifier nucleic acid sequence, and evaluating based on measurement of the extension reaction, the presence, absence, or relative abundance of one or more codons.

The extension reaction requires a primer, a polymerase as well as a collection of deoxyribonucleotide triphosphates (abbreviated dNTP's herein) to proceed. An extension product may be obtained in the event the primer is sufficient complementary to an identifier oligonucleotide for a polymerase to recognise the double helix as a substrate. After binding of the polymerase to the double helix, the deoxyribonucleotide triphosphates (blend of dATP, dCTP, dGTP, and dTTP) are incorporated into the extension product using the identifier oligonucleotide as identifier. The conditions allowing for the extension reaction to occur usually includes a suitable buffer. The buffer may be any aqueous or organic solvent or mixture of solvents in which the polymerase has a sufficient activity. To facilitate the extension process the polymerase and the mixture of dNTP's are generally included in a buffer which is added to the identifier oligonucleotide and primer mixture. An exemplary kit comprising the polymerase and the nNTP's for performing the extension process comprises the following: 50 mM KCl; 10 mM Tris-HCl at pH 8.3; 1.5 mM MgCl2 ; 0.001% (wt/vol)

gelatin, 200 µM dATP; 200 µM dTTP; 200 µM dCTP; 200 µM dGTP; and 2.5 units Thermus aquaticus (Taq) DNA polymerase I (U.S. Pat. No. 4,889,818) per 100 microliters (µl) of buffer.

The primer may be selected to be complementary to one or more codons or parts of such codons. The length of the primers may be determined by the length of the codons, however, the primers usually are at least about 11 nucleotides in length, more preferred at least 15 nucleotides in length to allow for an efficient extension by the polymerase. The presence or absence of one or more codons is indicated by the presence of or absence of an extension product. The extension product may be measured by any suitable method, such as size fractioning on an agarose gel and staining with ethidium bromide.

In a preferred embodiment the admixture of identifier oligonucleotide and primer is termocycled to obtain a sufficient number of copies of the extension product. The thermocycling is typically carried out by repeatedly increasing and decreasing the temperature of the mixture within a temperature range whose lower limit is about 30 degrees Celsius (30°C) to about 55°C and whose upper limit is about 90°C to about 100° C. The increasing and decreasing can be continuous, but is preferably phasic with time periods of relative temperature stability at each of temperatures favouring polynucleotide synthesis, denaturation and hybridization.

When a single complex is analysed in accordance with the present method, the result may be used to verify the presence or absence of a specific chemical entity during the formation of the display molecule. The formation of an extension product is indicative of the presence of an oligonucleotide part complementary to the primer in the identifier oligonucleotide. Conversely, the absence of an extension product is indicative of the absence of an oligonucleotide part complementary to the primer in the identifier oligonucleotide. Selecting the sequence of the primer such that it is complementary to one or more codons will therefore provide information of the structure of the encoded molecule coded for by this codon(s).

In a preferred aspect of the invention, in the mixture of the identifier oligonucleotide and the primer oligonucleotide, a second primer complementary to a sequence of the extension product is included. The second primer is also termed reverse primer

and ensures an exponential increase of the number of produced extension products. The method using a forward and reverse primer is well known to skilled person in the art and is generally referred to as polymerase chain reaction (abbreviated PCR) in the present application with claims. In one embodiment of the invention the reverse primer is annealed to a part of the extension product downstream, i.e. near the 3´end of the extension product, or a part complementing the coding part of the identifier oligonucleotide. In another embodiment, the first primer (forward primer) anneals to an upstream position of the identifier oligonucleotide, preferably before the coding part, and the reverse primer anneals to a sequence of the extension product complementing one or more codons or parts thereof.

The amplicons resulting from the PCR process may be stained during or following the reaction to ease the detection. A staining after the PCR process may be prepared with e.g. ethidium bromide or a similar staining agent. As an example, amplicons from the PCR process is run on an agarose gel and subsequently stained with ethidium bromide. Under UV illumination bands of amplicons becomes visible. It is possible to incorporate the staining agent in the agarose gel or to allow a solution of the staining agent to migrate through the gel. The amplicons may also be stained during the PCR process by an intercalating agent, like CYBR. In presence of the intercalating agent while the amplification proceeds it will incorporate in the double helix. The intercalation agent may then be made visible by irradiation by a suitable source.

The intensity of the staining is informative of the relative abundance of a specific amplicon. Thus, it is possible to quantify the occurrence of a codon in an identifier oligonucleotide. When a library of bifunctional complexes has been subjected to a selection the codons in the pool of identifier oligonucleotides which has been selected can be quantified using this method. As an example a sample of the selected identifier oligonucleotides is subjected to various PCR amplifications with different primers in separate compartments and the PCR product of each compartment is analysed by electrophoresis in the presence of ethidium bromide. The bands that appear can be quantified by a densitometric analysis after irradiation by ultraviolet light and the relative abundance of the codons can be measured.

Alternatively, the primers may be labelled with a suitable small molecule, like biotin or digoxigenin. A PCR-ELISA analysis may subsequently be performed based on the amplicons comprising the small molecule. A preferred method includes the application of a solid support covered with streptavidin or avidin when biotin is used as label and anti-digoxigenin when digoxigenin is used as the label. Once captured, the amplicons can be detected using an enzyme-labelled avidin or anti-dixigenin reporter molecule similar to a standard ELISA format.

To avoid laborious post-PCR handling steps required to evaluate the amplicons, it is in a certain embodiment preferred to measure the extension process "real time". Several real time PCR processes has been developed and all the suitable real time PCR process available to the skilled person in the art can be used in the evaluating step of the present invention and are include in the present scope of protection. The PCR reactions discussed below are of particular interest.

The monitoring of accumulating amplicons in real time has been made possible by labelling of primers, probes, or amplicons with fluorogenic molecules. The real time PCR amplification is usually performed with a speed faster than the conventional PCR, mainly due to reduced cycles time and the use of sensitive methods for detection of emissions from the fluorogenic labels. The most commonly used fluorogenic oligoprobes rely upon fluorescent resonance energy transfer (FRET) between fluorogenic labels or between one flourophor and a dark or "black-hole" non-fluorescent quencher (NFQ), which disperse energy as heat rather than fluorescence. FRET is a spectroscopic process by which energy is passed between molecules separated by 10-100 Å that have overlapping emission and absorption spectra. An advantage of many real time PCR methods is that they can be carried out in a closed system, i.e. a system which does not need to be opened to examine the result of the PCR. A closed system implies a reduced result turnaround, minimisation of the potential for carry-over contamination and the ability to closely scrutinise the essay's performance.

The real time PCR methods currently available to the skilled person can be classified into either amplicon sequence specific or non-specific methods. The basis for the non-specific detection methods is a DNA-binding fluorogenic molecule. Included in this class are the earliest and simplest approaches to real time PCR. Ethidium

bromide, YO-PRO-1, and SYBR® green 1 all fluorescence when associated with double stranded DNA which is exposed to a suitable wavelength of light. This approach requires the fluorescent agent to be present during the PCR process and provides for a real time detection of the fluorescent agent as it is incorporated into the double stranded helix.

The amplicons sequence specific methods includes, but are not limited to, the TaqMan®, hairpin, LightCycler®, Sunrise®, and Scorpion® methods. The LightCycler® method also designated "HybProbes" make use of a pair of adjacent, fluorogenic hybridisation oligonucleotide probes. A first, usually the upstream oligoprobe is labelled with a 3' donor fluorophore and the second, usually the downstream probe is commonly labelled with either a Light cycler Red 640 or Red 705 acceptor fluorophore a the 5' terminus so that when both oligoprobes are hybridised the two fluorophores are located in close proximity, such as within 10 nm, of each other. The close proximity provides for the emission of a fluorescence when irradiated with a suitable light source, such a blue diode in case of the LightCycler®. The region for annealing of the probes may be any suitable position that does not interfere with the primer annealing. In a suitable setup, the site for binding the probes are positioned downstream of the codon region on the identifier oligonucleotide. Alternatively, when a reverse primer is used, the region for annealing the probes may be at the 3' end of the strand complementing the identifier oligonucleotide. Another embodiment of the LightCycler method includes that the pair of oligonucleotide probes are annealed to one or more codons and primer sites exterior to the coding part of the identifier oligonucleotide are used for PCR amplification.

The TaqMan® method, also referred to as the 5' nuclease or hydrolysis method, requires an oligoprobe, which is attached to a reporter flourophor, such as 6-carboxy-fluoroscein, and a quencher fluorophore, such as 6-carboxy-tetramethyl-rhodamine, at each end. When in close proximity, i.e. annealed to an identifier oligonucleotide, or a sequence complementing the identifier oligonucleotide, the quencher will "hijack" the emissions that have resulted from the excitation of the reporter. As the polymerase progresses along the relevant strand, it displaces and the hydrolyses the oligoprobe via its 5'→3' endonuclease activity. Once the reporter is removed from the extinguishing influence of the quencher, it is able to release excitation energy at a wavelength that can be monitored by a suitable instrument,

such as ABI Prism® 7700. The fractional cycle number at which the real-time fluorescence signal mirrors progression of the reaction above the background noise is normally used as an indicator of successful identifier oligonucleotide amplification. This threshold cycle (C_T) is defined as the PCR cycle in which the gain in fluorescence generated by the accumulating amplicons exceeds 10 standard deviations of the mean base line fluorescence. The C_T is proportional to the number of identifier oligonucleotide copies present in the sample. The TaqMan probe is usually designed to hybridise at a position downstream of a primer binding site; be it a forward or a reverse primer. When the primer is designed to anneal to one or more codons of the identifier oligonucleotide, the presence of these one or more codons is indicated by the emittance of light. Furthermore, the quantity of the identifier oligonucleotides comprising the one or more codons may be measured by the C_T value.

The Hairpin method involves an oligoprobe, in which a fluorophore and a quencher are positioned at the termini. The labels are hold in close proximity by distal stem regions of homologous base pairing deliberately designed to create a hairpin structure which result in quenching either by FRET or a direct energy transfer by a collisional mechanism due to the intimate proximity of the labels. When direct energy transfer by a collision mechanism is used the quencher is usually different from the FRET mechanism, and is suitably 4-(4'-dimethylamino-phenylazo)-benzene (DABCYL). In the presence of a complementary sequence, usually downstream of a primer, or within the bounds of the primer binding sides in case of more than one a single primer, the oligoprobe will hybridise, shifting into an open configuration. The fluorophore is now spatially removed from the quencher's influence and fluorescence emissions are monitored during each cycle. In a certain aspect, the hairpin probe may be designed to anneal to a codon in order to detect this codon if present on the identifier oligonucleotide. This embodiment may be suitable if codons only differs from each other with a single or a few nucleotides, because is in well-known that the occurrence of a mismatch between a hairpin oligoprobe and its target sequence has a greater destabilising effect on the duplex than the introduction of an equivalent mismatch between the target oligonucleotide and a linear oligoprobe. This is probably because the hairpin structure provides a highly stable alternate conformation.

The Sunrise and Scorpion methods are similar in concept to the hairpin oligoprobe, except that the label becomes irreversible incorporated in to the PCR product. The Sunrise method involves a primer (commercially available as Amplifluor™ hairpin primers) comprising a 5' fluorophore and a quencher, e.g. DABCYL. The labels are separated by complementary stretches of sequence that create a stem when the sunrise primer is closed. At the 3' terminus is a target specific primer sequence. In a preferred embodiment the target sequence is a codon, optionally more codons. The sunrise primer's sequence is intended to be duplicated by the nascent complementary stand and, in this way, the stem is destabilised, the two fluorophores are held apart, usually between 15 and 25 nucleotides, and the fluorophore is free to emit its excitation energy for monitoring. The Scorpion primer resembles the sunrise primer, but derivate in having a molety that blocks duplication on the signalling portion of the scorpion primer. The blocking moiety is typically hexethylene glycol. In addition to the difference in structure, the function of the scorpion primers differs slightly in that the 5' region of the oligonucleotide is designed to hybridise to a complementary region within the amplicons. In a certain embodiment the complementary region is a codon on the identifier oligonucleotide. The hybridisation forces the labels apart disrupting the hairpin and permitting emission in the same way as the hairpin probes.

After the selection has been performed the codon profile is indicative of the chemical entities that have been used in the synthesis of encoded molecules having a certain property, such as an affinity towards a target. In the event the selection has been sufficient effective it may be possible directly to deduce a part or the entire structure of encoded molecules with the desired property. Alternatively, it may be possible to deduce a structural unit appearing more frequently among the encoded molecules after the selection, which gives important information to the structure-activity-relationship (SAR). If the selection process has not narrowed the size of the library to a manageable number, the formation of a second-generation library is useful. In the formation of the second-generation library chemical entities, which have not been involved in the synthesis of encoded molecules that have been successful in the selection may be omitted, thus limiting the size of the new library and at the same time increasing the concentration of complexes with the requested property, e.g. the ability to bind to a target. The second-generation library may then be subjected to more stringent selection conditions to allow only the encoded molecules with a higher affinity to bind to the target. The second-generation library may also be

generated using the chemical entities coded for in addition to certain chemical entities suspected of increasing the performance of the final encoded molecule. The indication of certain successful chemical entities may be obtained from the SAR. The use in a second-generation library of chemical entities, which have proved to be interesting for further investigation in a preceding library, may thus entail a shuffling with new chemical entities that may focus the second-generation library in a certain desired direction.

An Example of implicit identification of codons includes that the nucleic acid fragment is associated with a chemical entity precursor capable of being transferred to a recipient reactive group. The recipient reactive group may be a part of a chemical scaffold and the chemical entity precursor may add a structural unit to said scaffold. It is preferred that the nucleic acid fragment codes for the chemical entity. In some aspects of the present invention each member of the nucleic acid fragment pool comprises an anticodon, which identifies the chemical entity. When a plurality of chemical entities are present the anticodon is preferably unique, i.e. a unique correspondence between the chemical entities and the associated anticodons exists.

The identifier nucleic acid sequence comprises codons, which may be able to pair with one or more anticodons of the pool of nucleic acid fragments. The pairing between one or more codons of an identifier nucleic acid sequence and one or more anticodons is preferably specific, i.e. the one or more codons of the identifier nucleic acid sequence are only recognized by particular anticodons. The nucleic acid fragment containing more than one anticodon can encode for scaffold molecules where each anticodon encodes for specific chemical entities of that scaffold molecule. The specific pairing makes it possible implicitly to decode the codon of an identifier nucleic acid sequence. In the method according to the invention, non-specific pairing between codons and anticodons can be cleaved with an enzyme or chemically treated to break the double stranded nucleotides. The non-pairing region can be cleaved using enzymes that cleaves specifically nucleotide sequences with mismatches. Notably, the enzyme is selected from T4 endonuclease VII, T4 endonuclease I, CEL I, nuclease S1, or variants thereof. The cleavage is preferable used when more than one codon and anticodon is involved in pairing between the identifier nucleic acid sequence and the nucleic acid fragment.

The pool of nucleic acid fragments associated with a chemical entity may comprise anticodons complemented by codons of one or more identifier nucleic acid sequence as well as anticodons which are not complemented by codons on any identifier nucleic acid sequence. In other words, the amount of genetic information contained in the anticodons of the pool is larger than the amount of genetic information complemented by the codons.

The contacting of the one or more identifier nucleic acid sequences with the pool of nucleic acid fragments are usually conducted at conditions, which allow for hybridisation, i.e. conditions at which cognate nucleic acid sequences can anneal to each other. To facilitate the recovery of nucleic acid fragments, which have annealed to the identifier nucleic acid sequences, the identifier nucleic acid sequences are usually immobilized on a solid support. Examples of suitable solid supports include beads and column material, e.g. beads and column material associated with a second part of the affinity pair to bind identifier nucleic acid sequences attached to the first part of the molecular affinity pair. In certain aspects of the invention the solid support is associated with streptavidin and the identifier nucleic acid sequences are attached to biotin.

When the identifier nucleic an acid sequences are immobilized on a solid support the pool of nucleic acid fragments is typically present in a mobile phase, i.e. dissolved in a liquid. The identifier nucleic acids will hybridise to these nucleic acid fragments in the pool which are sufficient complementary to a particular part of an identifier nucleic acid sequence for a binding to occur. Fragments not finding any complementing sequence will remain in the solution. In the event, the identifier nucleic acid sequences are segregated into codons and the fragments comprises anticodons, the anticodons which are able to anneal to a codons will be caught while fragments not having a cognate codon will be maintained in the mobile phase. When codons and anticodons are present in the method of the present invention, specific hybridisation implies that the tendency of an anticodon to cross-hybridise to another codon will be impede or avoided. To avoid cross-hybridisation, codons may be designed such that each codon is distinguished from all other codons be one, two or more mismatching nucleotides.

The mobile phase is subsequently separated from the solid phase e.g. by washing, and the enriched pool of fragments is recovered. The recovery of the nucleic acid fragments are usually done by subjecting the hybrid to denaturing conditions, i.e. conditions which separate the two strands. If the parent nucleic acid sequences are immobilized on beads, the separation of the fragments can be effected using denaturing conditions and centrifugation/spinning.

The enriched pool of nucleic acid fragments associated with a chemical entity may be used directly to prepare a next generation library of complexes, in which each member of the library comprises an encoded molecule and the nucleic acid sequence which codes for this molecule. In one embodiment of the invention, building blocks comprising a particular transferable chemical entity associated with an anticodon corresponding to the anticodons of the detected fragments are used in the generation of the next generation library. In another embodiment, additional building blocks are added having modified transferable chemical entities in order to improve on a certain property of the encoded molecule.

The complexes may be prepared by various known methods starting from the nucleic acid fragment comprising the anticodon and the chemical entity, as disclosed above. According to a particular method, the next generation library is formed by a) mixing under hybridisation conditions, nascent bifunctional complexes comprising a chemical entity or a reaction product of chemical entities, and an identifier nucleic acid sequence comprising codon(s) identifying said chemical entities, with the recovered nucleic acid fragments, said fragments comprising an oligonucleotide sufficient complementary to at least a part of the identifier nucleic acid sequence to allow for hybridisation, a transferable chemical entity and an anticodon identifying the chemical entity, to form hybridisation products; and b) transferring the chemical entities of the nucleic acid fragments to the nascent bifunctional complexes through a reaction involving a reactive group of the nascent bifunctional complex, in conjunction with a transfer of the genetic information of the anticodon.

Preferably, the above method for preparing the next generation library comprises the further step of c) separating the components of the hybridisation product and recovering the complexes. If further chemical entities are intended to participate in

the formation of the encoded molecule of the nascent complex, steps a) through c) are repeated as appropriate using the recovered complexes in step c) as the nascent bifunctional complexes in step a) of the next round.

The genetic information of the anticodon may be transferred to the nascent complex by a variety of methods. According to a first embodiment the genetic information of the anticodon is transferred by enzymatically extending the oligonucleotide identifier region to obtain a codon attached to the bifunctional complex having received the chemical entity. A second embodiment implies that genetic information of the anticodon is transferred to the nascent complexes by hybridisation to a cognate codon of the nascent complex.

According to the first embodiment, the enriched pool of fragments comprises an affinity oligonucleotide sufficient complementary to an identifier region of the nascent complex, said oligonucleotide being distinct from the anticodon. Accordingly, the oligonucleotide identifier region of the nascent complex anneals to the affinity oligonucleotide of the building block to form the hybridisation product, while the anticodon remains single stranded. Subsequently, the chemical entity is transferred to the recipient reactive group of the complex to form the encoded molecule prior to, simultaneously with, or subsequent to the enzymatically extension of the hybridisation product using the anticodon as identifier. Specific examples of suitable enzymes are polymerases and ligases, which requires dNTPs and oligonucleotides, respectively as substrates. The method for forming the complexes according to this first embodiment is the subject PCT/DK03/00739, the content thereof being incorporated herein by reference.

According to the second embodiment, the anticodon form part of the affinity oligonucleotide, i.e. the anticodon is a part of or the entire affinity oligonucleotide. Initially, a plurality of identifiers comprising different codons and/or different order of codons is provided. The identifiers are associated with a recipient reactive group, i.e. the reactive group may be covalently attached to the identifier or attached by hybridisation. Notably, a codon of the identifier may be used for the attachment of a building block harbouring the reactive group. The identifiers are subsequently contacted with the enriched pool of building blocks, i.e. nucleic acid fragments associated with a transferable chemical entity. The mixture of identifiers and building blocks are maintained

at hybridisation conditions to anneal the anticodon of the building blocks to the cognate codon of the identifier. After or simultaneously with the annealing step, the chemical entity is transferred to the recipient reactive group of the identifier. The method for forming the complexes according to the second embodiment is the subject of various patent applications, including WO 02/103008, WO 02/074929, Danish patent application No. PA 2002 01347, and US provisional patent application No. 60/409,968. The content of these patent applications are incorporated herein by reference in their entirety.

The new generation of library complexes may be used in a partition step, in which the library of complexes is subjected to a condition partitioning complexes displaying a predetermined property from the remainder of the next generation library, as explained above. Thus, using the present method, it is possible to repeat the partitioning procedure a desired number of times using still more stringent conditions, until a single or a few encoded molecules are identified which display the desired property to a high extent. When the partitioning is based on an affinity assay, the library of encoded molecules are increasingly narrowed in size from one generation to the next and at the same time the high affinity binders are increased in concentration.

The outcome of a codon analysis will be dependent of the enrichment factor in the selection process. An efficient and specific selection will generate a large difference between the specific binders compared to the background. Still, there will be a large amount of molecules in the background that will reduce the possibility to obtain measurable differences between the binders and the background in the codon analysis procedure. If the enrichment factor (or too large library) is not good enough to distinguish a specific binder among the background binders, the signal in the codon analysis will probably not be detectable. However, there will be a continuing of binders that use a certain chemical entity in a certain position. These "non optimal" binders (a certain important chemical entity in one position and less important in the other position) will be many due to the diversity obtained when only one (or a few) positions are important in the selection process. Therefore, the sum of all molecules with a preferable chemical entity in a certain position will be larger than the sum of all molecules with a non-binding chemical entity, which will make the codon analysis easier.

This invention may involve an extensive analysis of all the chemical entities in a library and how they are involved in the binding to targets. This information can be used both to design new libraries and in the final process where the lead structures are produced and pre-clinical candidates are picked. The extensive data obtained in the codon analysis can for instance be used for selecting candidates with the appropriate specificity. This can be done if selection has been performed on a family of proteins where one of the members is the target.

The invention enables pharmacophore identification and transformation into small molecule drugs. In cases where peptide-like libraries is used, the peptide/petidomimetic lead to small molecule conversion process is supported by medicinal chemistry and cheminformatics and guided by matching the pharmacophore derived from massive structure activity relationship (SAR) data information from the codon analysis. A "pharmacophore" is a description of the structural criteria a molecule must fulfil in order that it is active against a specified biological receptor. These criteria are usually the 3D spatial relationships of a set of chemical features, and sometimes include the steric boundaries, within which the molecule must fit. There is a set of software methods, which automatically infers such pharmacophores, given a SAR, in the absence of direct macromolecular structural data.

The extensive SAR information obtained using the codon analyses described in this invention can be combined with molecular modeling technologies to refine for example pharmacophore models and the plausible interactions between the potential binders and a target.

The codon analysis is also a valuable experimental tool for SAR on weak binders. The codon analysis measures the abundance of chemical entities after a selection in all binding molecules. Thus, even week binders, which there might be many of, is detected even though the detected codon is selected in many different combinations. The selection procedure can also be tuned to enrich predominately for weak binders, which will simplify the codon analysis data.

This invention is also suitable for replacing the laborious task of extracting SAR information by hand with an automated process using suitable algorithm and software programs. The codon analysis (e.g. array or QPCR measurements) can be directly

feed into a data handling software program that use both the codon abundances and structural data to generate SAR information and potential pharmacophore models.

The SAR information and potential pharmacophore models obtained from the codon analysis can be used to design focused libraries in an array format allowing massive and parallel testing. Thus, the selection procedure and codon analysis can be seen as a diversity reduction step to allow a complete test of potential binders in an array format.

Various methods for identifying the codons of the identifiers of step iii) are disclosed herein. When a pool of partitioned identifier nucleic acid sequences is subjected to the identification step it is normally not practically to decode a sufficient number of sequences comprising the entire "genome" of an encoded molecule to ensure that all interesting encoded molecules have been revealed. Therefore, a modified sequencing technique preferably identifies the codons in each position occurring with the highest frequency. The next generation library is then build using in each position the chemical entities occurring with the highest frequency.

In a certain embodiment of the invention, the codon identification step uses the entire population of identifier nucleic acid sequences in the analysis and informs the experimenter of the relative abundance of each codon in a certain position. The codon information may be obtained using microarray, QPCR, or any equivalent method for revealing the identity of codons. In contrary, sequencing a subset of identifier nucleic acid sequences only provides the experimenter with a limited insight as to the population of codons and the corresponding encoded molecules.

Detailed Description of the Invention

Complex

The complex comprises an encoded molecule and an identifier oligonucleotide. The identifier comprises codons that identify the encoded molecule. Preferably, the identifier oligonucleotide identifies the encoded molecule uniquely, i.e. in a library of complexes a particular identifier is capable of distinguishing the molecule it is attached to from the rest of the molecules.

The encoded molecule and the identifier may be attached directly to each other or through a bridging moiety. In one aspect of the invention, the bridging moiety is a selectively cleavable linkage.

The identifier oligonucleotide may comprise two or more codons. In a preferred aspect the identifier oligonucleotide comprises three or more codons. The sequence of each codon can be decoded utilizing the present method to identify reactants used in the formation of the encoded molecule. When the identifier comprises more than one codon, each member of a pool of chemical entities can be identified and the order of codons is informative of the synthesis step each member has been incorporated in.

In a certain embodiment, the same codon is used to code for several different chemical entities. In a subsequent identification step, the structure of the encoded molecule can be deduced taking advantage of the knowledge of different attachment chemistries, steric hindrance, deprotection of orthogonal protection groups, etc. In another embodiment, the same codon is used for a group of chemical entities having a common property, such as a lipophilic nature, a certain attachment chemistry etc. In a preferred embodiment, however, the codon is unique i.e. a similar combination of nucleotides does not appear on the identifier oligonucleotide coding for another chemical entity. In a practical approach, for a specific chemical entity, only a single combination of nucleotides is used. In some aspects of the invention, it may be advantageous to use several codons for the same chemical entity, much in the same way as Nature uses up to six different codons for a single amino acid. The two or more codons identifying the same chemical entity may carry further information related to different reaction conditions.

The sequence of the nucleotides in each codon may have any suitable length. The codon may be a single nucleotide or a plurality of nucleotides. In some aspects of the invention, it is preferred that each codon independently comprises four or more nucleotides, more preferred 4 to 30 nucleotides. In some aspects of the invention the lengths of the codons vary.

A certain codon may be distinguished from any other codon in the library by only a single nucleotide. However, to facilitate a subsequent decoding process and to in-

crease the ability of the primer to discriminate between codons it is in general desired to have two or more mismatches between a particular codon and any other codon appearing on identifier oligonucleotide. As an example, if a codon length of 5 nucleotides is selected, more than 100 nucleotide combinations exist in which two or more mismatches appear. For a certain number of nucleotides in the codon, it is generally desired to optimize the number of mismatches between a particular codon relative to any other codon appearing in the library.

The identifier oligonucleotide will in general have at least two codons arranged in sequence, i.e. next to each other. Two neighbouring codons may be separated by a framing sequence. Depending on the encoded molecule formed, the identifier may comprise further codons, such as 3, 4, 5, or more codons. Each of the further codons may be separated by a suitable framing sequence. Preferably, all or at least a majority of the codons of the identifier are separated from a neighbouring codon by a framing sequence. The framing sequence may have any suitable number of nucleotides, e.g. 1 to 20. Alternatively, codons on the identifier may be designed with overlapping sequences.

The framing sequence, if present, may serve various purposes. In one setup of the invention, the framing sequence identifies the position of the codon. Usually, the framing sequence either upstream or downstream of a codon comprises information which positions the chemical entity and the reaction conditions in the synthesis history of the encoded molecule. The framing sequence may also or in addition provide for a region of high affinity. The high affinity region may ensure that a hybridisation event with an anti-codon will occur in frame. Moreover, the framing sequence may adjust the annealing temperature to a desired level.

A framing sequence with high affinity can be provided by incorporation of one or more nucleobases forming three hydrogen bonds to a cognate nucleobase. Examples of nucleobases having this property are guanine and cytosine. Alternatively, or in addition, the framing sequence may be subjected to backbone modification. Several back bone modifications provides for higher affinity, such as 2'-O-methyl substitution of the ribose moiety, peptide nucleic acids (PNA), and 2'-4' O-methylene cyclisation of the ribose moiety, also referred to as LNA (Locked Nucleic Acid).

The sequence comprising a codon and an adjacent framing sequence has in a certain aspect of the invention a total length of 11 nucleotides or more, preferably 15 nucleotides or more. A primer may be designed to complementary to the codon sequence as well as the framing sequence. The presence of an extension reaction under conditions allowing for such reaction to occur is indicative of the presence of the chemical entity encoded in the codon as well as the position said chemical entity has in the entire synthesis history of the encoded molecule.

The identifier may comprise flanking regions around the coding section. The flanking regions can also serve as priming sites for amplification reactions, such as PCR or as binding region for oligonucleotide probe. The identifier may in certain embodiments comprise an affinity region having the property of being able to hybridise to a building block.

It is to be understood that when the term identifier oligonucleotide is used in the present description and claims, the identifier oligonucleotide may be in the sense or the anti-sense format, i.e. the identifier can be a sequence of codons which actually codes for the encoded molecule or can be a sequence complementary thereto. Moreover, the identifier may be single-stranded or double-stranded, as appropriate.

The encoded molecule part of the complex is generally of a structure expected of having an effect according to the property sought for, e.g. the encoded molecule has a binding affinity towards a target. When the target is of pharmaceutical importance, the encoded molecule is generally a possible drug candidate. The complex may be formed by tagging a library of different possible drug candidates with a tag, e.g. a nucleic acid tag identifying each possible drug candidate. In another embodiment of the invention, the molecule formed by a variety of reactants which have reacted with each other and/or a scaffold molecule. Optionally, this reaction product may be post-modified to obtain the final molecule displayed on the complex. The post-modification may involve the cleavage of one or more chemical bonds attaching the encoded molecule to the identifier in order more efficiently to display the encoded molecule.

The formation of an encoded molecule generally starts by a scaffold, i.e. a chemical unit having one or more reactive groups capable of forming a connection to another

reactive group positioned on a chemical entity, thereby generating an addition to the original scaffold. A second chemical entity may react with a reactive group also appearing on the original scaffold or a reactive group incorporated by the first chemical entity. Further chemical entities may be involved in the formation of the final reaction product. The formation of a connection between the chemical entity and the nascent encoded molecule may be mediated by a bridging molecule. As an example, if the nascent encoded molecule and the chemical entity both comprise an amine group a connection between these can be mediated by a dicarboxylic acid. A synthetic molecule is in general produced in vitro and may be a naturally occurring or an artificial substance. Usually, a synthetic molecule is not produced using the naturally translation system in an in vitro process.

The chemical entities that are precursors for structural additions or eliminations of the encoded molecule may be attached to a building block prior to the participation in the formation of the reaction product leading the final encoded molecule. Besides the chemical entity, the building block generally comprises an anti-codon. In some embodiments the building blocks also comprise an affinity region providing for affinity towards the nascent complex.

Thus, the chemical entities are suitably mediated to the nascent encoded molecule by a building block, which further comprises an anticodon. The anti-codon serves the function of transferring the genetic information of the building block in conjunction with the transfer of a chemical entity. The transfer of genetic information and chemical entity may occur in any order. The chemical entities are preferably reacted without enzymatic interaction in some aspects of the invention. Notably, the reaction of the chemical entities is preferably not mediated by ribosomes or enzymes having similar activity. In other aspects of the invention, enzymes are used to mediate the reaction between a chemical entity and a nascent encoded molecule.

According to certain aspects of the invention the genetic information of the anti-codon is transferred by specific hybridisation to a codon on a nucleic acid identifier. Another method for transferring the genetic information of the anti-codon to the nascent complex is to anneal an oligonucleotide complementary to the anti-codon and attach this oligonucleotide to the complex, e.g. by ligation. A still further method

involves transferring the genetic information of the anti-codon to the nascent complex by an extension reaction using a polymerase and a mixture of dNTPs.

The chemical entity of the building block may in most cases be regarded as a precursor for the structural entity eventually incorporated into the encoded molecule. In other cases the chemical entity provides for the eliminations of chemical units of the nascent encoded molecule. Therefore, when it in the present application with claims is stated that a chemical entity is transferred to a nascent encoded molecule it is to be understood that not necessarily all the atoms of the original chemical entity is to be found in the eventually formed encoded molecule. Also, as a consequence of the reactions involved in the connection, the structure of the chemical entity can be changed when it appears on the nascent encoded molecule. Especially, the cleavage resulting in the release of the entity may generate a reactive group which in a subsequent step can participate in the formation of a connection between a nascent complex and a chemical entity.

The chemical entity of the building block comprises at least one reactive group capable of participating in a reaction which results in a connection between the chemical entity of the building block and another chemical entity or a scaffold associated with the nascent complex. The number of reactive groups which appear on the chemical entity is suitably one to ten. A building block featuring only one reactive group is used i.a. in the end positions of polymers or scaffolds, whereas building blocks having two reactive groups are suitable for the formation of the body part of a polymer or scaffolds capable of being reacted further. One, two or more reactive groups intended for the formation of connections, are typically present on scaffolds. Non-limiting examples of scaffolds are oplates, steroids, benzodiazepines, hydantoines, and peptidylphosphonates.

The reactive group of the chemical entity may be capable of forming a direct connection to a reactive group of the nascent complex or the reactive group of the building block may be capable of forming a connection to a reactive group of the nascent complex through a bridging fill-in group. It is to be understood that not all the atoms of a reactive group are necessarily maintained in the connection formed. Rather, the reactive groups are to be regarded as precursors for the structure of the connection.

The subsequent cleavage step to release the chemical entity from the building block can be performed in any appropriate way. In an aspect of the invention the cleavage involves usage of a chemical reagent or an enzyme. The cleavage results in a transfer of the chemical entity to the nascent encoded molecule or in a transfer of the nascent encoded molecule to the chemical entity of the building block. In some cases it may be advantageous to introduce new chemical groups as a consequence of linker cleavage. The new chemical groups may be used for further reaction in a subsequent cycle, either directly or after having been activated. In other cases it is desirable that no trace of the linker remains after the cleavage.

In another aspect, the connection and the cleavage is conducted as a simultaneous reaction, i.e. either the chemical entity of the building block or the nascent encoded molecule is a leaving group of the reaction. In some aspects of the invention, it is appropriate to design the system such that the connection and the cleavage occur simultaneously because this will reduce the number of steps and the complexity. The simultaneous connection and cleavage can also be designed such that either no trace of the linker remains or such that a new chemical group for further reaction is introduced, as described above.

The attachment of the chemical entity to the building block, optionally via a suitable spacer can be at any entity available for attachment, e.g. the chemical entity can be attached to a nucleobase or the backbone. In general, it is preferred to attach the chemical entity at the phosphor of the internucleoside linkage or at the nucleobase. When the nucleobase is used for attachment of the chemical entity, the attachment point is usually at the 7 position of the purines or 7-deaza-purins or at the 5 position of pyrimidines. The nucleotide may be distanced from the reactive group of the chemical entity by a spacer moiety. The spacer may be designed such that the conformational spaced sampled by the reactive group is optimized for a reaction with the reactive group of the nascent encoded molecule.

The encoded molecules may have any chemical structure. In a preferred aspect, the encoded molecule can be any compound that may be synthesized in a component-by-component fashion. In some aspects the synthetic molecule is a linear or branched polymer. In another aspect the synthetic molecule is a scaffolded molecule. The term "encoded molecule" also comprises naturally occurring

molecules like α-polypeptides etc, however produced *in vitro* usually in the absence of enzymes, like ribosomes. In certain aspects, the synthetic molecule of the library is a non-α-polypeptide.

The encoded molecule may have any molecular weight. However, in order to be orally available, it is in this case preferred that the synthetic molecule has a molecular weight less than 2000 Daltons, preferably less than 1000 Dalton, and more preferred less than 500 Daltons.

The size of the library may vary considerably pending on the expected result of the inventive method. In some aspects, it may be sufficient that the library comprises two, three, or four different complexes. However, in most events, more than two different complexes are desired to obtain a higher diversity. In some aspects, the library comprises 1,000 or more different complexes, more preferred 1,000,000 or more different complexes. The upper limit for the size of the library is only restricted by the size of the vessel in which the library is comprised. It may be calculated that a vial may comprise up to 10^{14} different complexes.

Methods for forming libraries of complexes

The encoded molecules associated with an identifier oligonucleotide having two or more codons that code for reactants that have reacted in the formation of the molecule part of the complex may be formed by a variety of processes. Generally, the preferred methods can be used for the formation of virtually any kind of encode molecule. Suitable examples of processes include prior art methods disclosed in WO 93/20242, WO 93/06121, WO 00/23458, WO 02/074929, and WO 02/103008, the content of which being incorporated herein by reference as well as methods of the present applicant not yet public available, including the methods disclosed in PCT/DK03/00739 filed 30 October 2003, and DK PA 2003 00430 filed 20 March 2003. Any of these methods may be used, and the entire content of the patent applications are included herein by reference.

Below five presently preferred embodiments are described. A first embodiment disclosed in more detail in WO 02/103008 is based on the use of a polymerase to incorporate unnatural nucleotides as building blocks. Initially, a plurality of identifier oligonucleotides is provided. Subsequently primers are annealed to each of the

identifiers and a polymerase is extending the primer using nucleotide derivatives, which have appended chemical entities. Subsequent to or simultaneously with the incorporation of the nucleotide derivatives, the chemical entities are reacted to form a reaction product. The encoded molecule may be post-modified by cleaving some of the linking moieties to better present the encoded molecule.

Several possible reaction approaches for the chemical entities are apparent. First, the nucleotide derivatives can be incorporated and the chemical entities subsequently polymerised. In the event the chemical entities each carry two reactive groups, the chemical entities can be attached to adjacent chemical entities by a reaction of these reactive groups. Exemplary of the reactive groups are amine and carboxylic acid, which upon reaction form an amide bond. Adjacent chemical entities can also be linked together using a linking or bridging moiety. Exemplary of this approach is the linking of two chemical entities each bearing an amine group by a bi-carboxylic acid. Yet another approach is the use of a reactive group between a chemical entity and the nucleotide building block, such as an ester or a hoister group. An adjacent building block having a reactive group such as an amine may cleave the interspaced reactive group to obtain a linkage to the chemical entity, e.g. by an amide linking group.

A second embodiment for obtainment of complexes disclosed in WO 02/103008 pertains to the use of hybridisation of building blocks to an identifier and reaction of chemical entities attached to the building blocks in order to obtain a reaction product. This approach comprises that identifiers are contacted with a plurality of building blocks, wherein each building block comprises an anti-codon and a chemical entity. The anti-codons are designed such that they recognise a sequence, i.e. a codon, on the identifier. Subsequent to the annealing of the anti-codon and the codon to each other a reaction of the chemical entity is effected.

The identifier may be associated with a scaffold. Building blocks bringing chemical entities in may be added sequentially or simultaneously and a reaction of the reactive group of the chemical entity may be effected at any time after the annealing of the building blocks to the identifier.

A third embodiment for the generation of a complex includes chemical or enzymatic ligation of building blocks when these are lined up on a identifier. Initially, identifiers are provided, each having one or more codons. The identifiers are contacted with building blocks comprising anti-codons linked to chemical entities. The two or more anti-codons annealed on an identifier are subsequently ligated to each other and a reaction of the chemical entities is effected to obtain a reaction product. The method is disclosed in more detail in DK PA 2003 00430 filed 20 March 2003.

A fourth embodiment makes use of the extension by a polymerase of an affinity sequence of the nascent complex to transfer the anti-codon of a building block to the nascent complex. The method implies that a nascent complex comprising a scaffold and an affinity region is annealed to a building block comprising a region complementary to the affinity section. Subsequently, the anti-codon region of the building block is transferred to the nascent complex by a polymerase. The transfer of the chemical entity may be transferred prior to, simultaneously with or subsequent to the transfer of the anti-codon. This method is disclosed in detail in PCT/DK03/00739.

A fifths embodiment also disclosed in PCT/DK03/00739 comprises reaction of a reactant with a reaction site on nascent bifunctional molecule and addition of a nucleic acid tag to the nascent bifunctional molecule using an enzyme, such as a ligase. When a library is formed, usually an array of compartments is used for reaction of reactants and enzymatic addition of tags with the nascent bifunctional molecule.

Thus, the codons are either pre-made into one or more identifiers before the encoded molecules are generated or the codons are transferred simultaneously with the formation of the encoded molecules.

After or simultaneously with the formation of the reaction product some of the linkers to the identifier may be cleaved, however, usually at least one linker is maintained to provide for the complex.

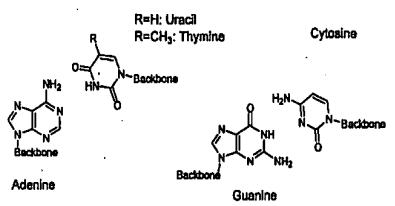
Nucleotides

The nucleotides used in the present invention may be linked together in a sequence of nucleotides, i.e. an oligonucleotide. Each nucleotide monomer is normally composed of two parts, namely a nucleobase moiety, and a backbone. The backbone may in some cases be subdivided into a sugar moiety and an internucleoside linker.

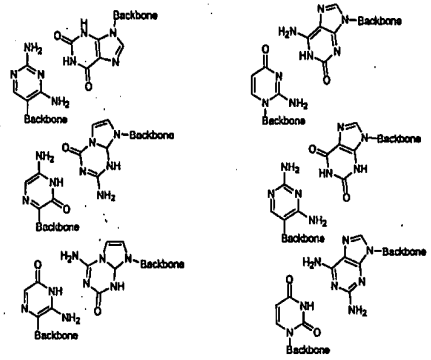
The nucleobase moiety may be selected among naturally occurring nucleobases as well as non-naturally occurring nucleobases. Thus, "nucleobase" includes not only the known purine and pyrimidine hetero-cycles, but also heterocyclic analogues and tautomers thereof. Illustrative examples of nucleobases are adenine, guanine, thymine, cytosine, uracil, purine, xanthine, diaminopurine, 8-oxo-N^6-methyladenine, 7-deazaxanthine, 7-deazaguanine, N^4,N^4-ethanocytosin, N^6,N^6-ethano-2,6-diaminopurine, 5-methylcytosine, 5-(C^3-C^6)-alkynylcytosine, 5-fluorouracil, 5-bromouracil, pseudoisocytosine, 2-hydroxy-5-methyl-4-triazolopyridine, isocytosine, isoguanine, inosine and the "non-naturally occurring" nucleobases described in Benner et al., U.S. Pat No. 5,432,272. The term "nucleobase" is intended to cover these examples as well as analogues and tautomers thereof. Especially interesting nucleobases are adenine, guanine, thymine, cytosine, 5-methylcytosine, and uracil, which are considered as the naturally occurring nucleobases in relation to therapeutic and diagnostic application in humans.

Examples of suitable specific pairs of nucleobases are shown below:

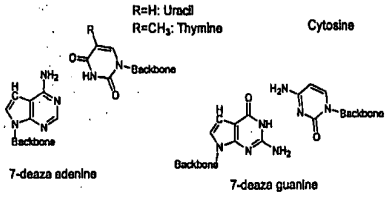
Natural Base Pairs



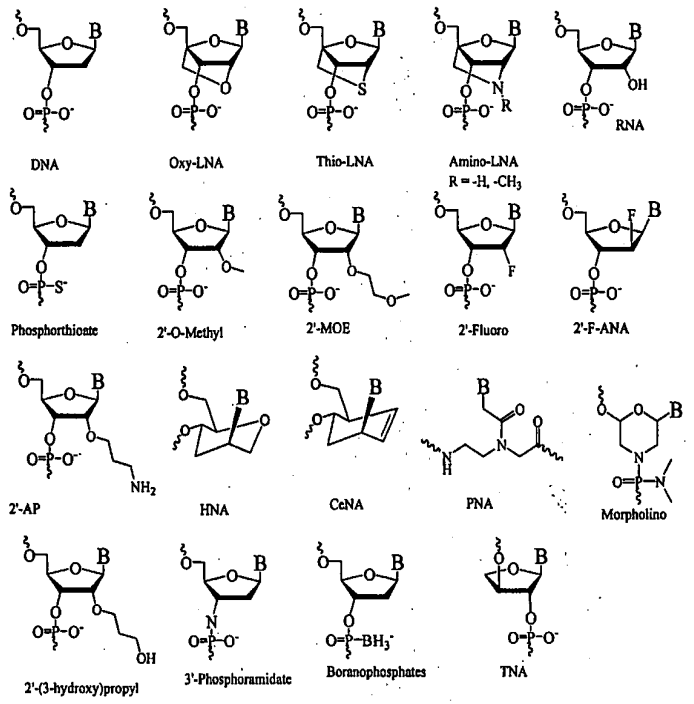
Synthetic Base Pairs



Synthetic purine bases pairing with natural pyrimidines



Suitable examples of backbone units are shown below (B denotes a nucleobase):

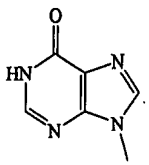


The sugar moiety of the backbone is suitably a pentose but may be the appropriate part of a PNA or a six-member ring. Suitable examples of possible pentoses include ribose, 2'-deoxyribose, 2'-O-methyl-ribose, 2'-flour-ribose, and 2'-4'-O-methylene-ribose (LNA). Suitably the nucleobase is attached to the 1' position of the pentose entity.

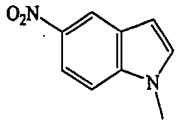
An internucleoside linker connects the 3' end of preceding monomer to a 5' end of a succeeding monomer when the sugar moiety of the backbone is a pentose, like ribose or 2-deoxyribose. The internucleoside linkage may be the natural occurring phospodiester linkage or a derivative thereof. Examples of such derivatives include phosphorothioate, methylphosphonate, phosphoramidate, phosphotriester, and phosphodithioate. Furthermore, the internucleoside linker can be any of a number of non-phosphorous-containing linkers known in the art.

Preferred nucleic acid monomers include naturally occurring nucleosides forming part of the DNA as well as the RNA family connected through phosphodiester linkages. The members of the DNA family include deoxyadenosine, deoxyguanosine, deoxythymidine, and deoxycytidine. The members of the RNA family include adenosine, guanosine, uridine, cytidine, and inosine. Inosine is a non-specific pairing nucleoside and may be used as universal base because inosine can pair nearly isoenergetically with A, T, and C. Other compounds having the same ability of non-specifically base-pairing with natural nucleobases have been formed. Suitable compounds which may be utilized in the present invention includes among others the compounds depicted below

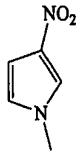
Examples of Universal Bases:



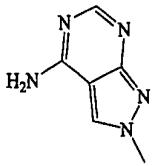
Inosine



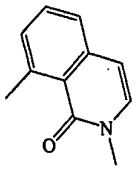
5-Nitroindole



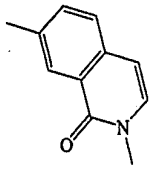
3-Nitropyrrole



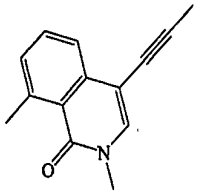
N^8-8aza-7deazaadenine



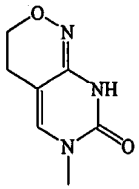
MICS



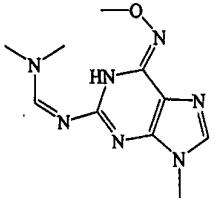
5MICS



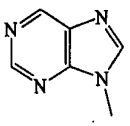
PIM



dP



dK



Nebularine

Building block

The chemical entities or reactants that are precursors for structural additions or eliminations of the encoded molecule may be attached to a building block prior to the participation in the formation of the reaction product leading to the final encoded molecule. Besides the chemical entity, the building block generally comprises an anti-codon.

The chemical entity of the building block comprises at least one reactive group capable of participating in a reaction, which results in a connection between the chemical entity of the building block and another chemical entity or a scaffold associated with the nascent complex. The connection is facilitated by one or more reactive groups of the chemical entity. The number of reactive groups, which appear on the chemical entity, is suitably one to ten. A building block featuring only one reactive group is used i.a. in the end positions of polymers or scaffolds, whereas building blocks having two reactive groups are suitable for the formation of the body part of a polymer or scaffolds capable of being reacted further. One, two or more reactive groups intended for the formation of connections are typically present on scaffolds.

The reactive group of the building block may be capable of forming a direct connection to a reactive group of the nascent complex or the reactive group of the building block may be capable of forming a connection to a reactive group of the nascent complex through a bridging fill-in group. It is to be understood that not all the atoms of a reactive group are necessarily maintained in the connection formed. Rather, the reactive groups are to be regarded as precursors for the structure of the connection.

The subsequent cleavage step to release the chemical entity from the building block can be performed in any appropriate way. In an aspect of the invention the cleavage involves usage of a reagent or an enzyme. The cleavage results in a transfer of the chemical entity to the nascent encoded molecule or in a transfer of the nascent encoded molecule to the chemical entity of the building block. In some cases it may be advantageous to introduce new chemical groups as a consequence of linker cleavage. The new chemical groups may be used for further reaction in a subsequent cycle, either directly or after having been activated. In other cases it is desirable that no trace of the linker remains after the cleavage.

In another aspect, the connection and the cleavage are conducted as a simultaneous reaction, i.e. either the chemical entity of the building block or the nascent encoded molecule is a leaving group of the reaction. In general, it is preferred to design the system such that the connection and the cleavage occur simultaneously because this will reduce the number of steps and the complexity. The simultaneous connection and cleavage can also be designed such that either no trace of the linker remains or such that a new chemical group for further reaction is introduced, as described above.

The attachment of the chemical entity to the building block, optionally via a suitable spacer can be at any entity available for attachment, e.g. the chemical entity can be attached to a nucleobase or the backbone. In general, it is preferred to attach the chemical entity at the phosphor of the internucleoside linkage or at the nucleobase. When the nucleobase is used for attachment of the chemical entity, the attachment point is usually at the 7 position of the purines or 7-deaza-purins or at the 5 position of pyrimidines. The nucleotide may be distanced from the reactive group of the chemical entity by a spacer moiety. The spacer may be designed such that the conformational space sampled by the reactive group is optimized for a reaction with the reactive group of the nascent encoded molecule or reactive site.

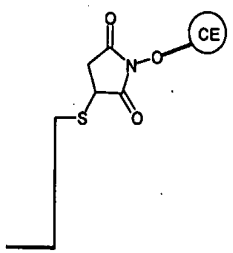
The anticodon complements the codon of the identifier oligonucleotide sequence and generally comprises the same number of nucleotides as the codon. The anticodon may be adjoined with a fixed sequence, such as a sequence complementing a framing sequence.

Various specific building blocks are envisaged. Building blocks of particular interest are shown below.

Building blocks transferring a chemical entity to a recipient nucleophilic group

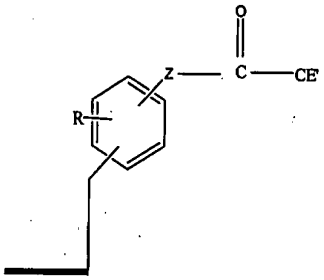
The building block indicated below is capable of transferring a chemical entity (CE) to a recipient nucleophilic group, typically an amine group. The bold lower horizontal line illustrates the building block comprising an anti-codon and the vertical line illustrates a spacer. The 5-membered substituted N-hydroxysuccinimid (NHS) ring

serves as an activator, i.e. a labile bond is formed between the oxygen atom connected to the NHS ring and the chemical entity. The labile bond may be cleaved by a nucleophilic group, e.g. positioned on a scaffold



The 5-membered substituted N-hydroxysuccinimid (NHS) ring serves as an activator, i.e. a labile bond is formed between the oxygen atom connected to the NHS ring and the chemical entity. The labile bond may be cleaved by a nucleophilic group, e.g. positioned on a scaffold, to transfer the chemical entity to the scaffold, thus converting the remainder of the fragment into a leaving group of the reaction. When the chemical entity is connected to the activator through a carbonyl group and the recipient group is an amine, the bond formed on the scaffold will an amide bond. The above building block is the subject of WO03078627A2, the content of which is incorporated herein in their entirety by reference.

Another building block, which may form an amide bond, is

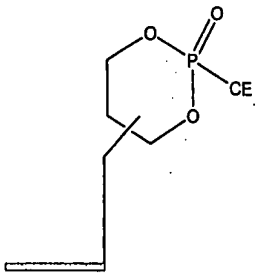


R may be absent or NO_2, CF_3, halogen, preferably Cl, Br, or I, and Z may be S or O. This type of building block is disclosed in WO03078626A2. The content of this patent application is incorporated herein in the entirety by reference.

A nucleophilic group can cleave the linkage between Z and the carbonyl group thereby transferring the chemical entity –(C=O)-CE' to said nucleophilic group.

Building blocks transferring a chemical entity to a recipient reactive group forming a C=C bond

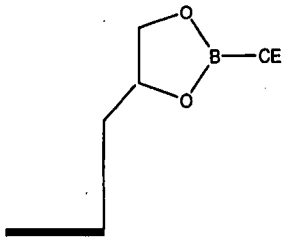
A building block as shown below is able to transfer the chemical entity to a recipient aldehylde group thereby forming a double bond between the carbon of the aldehyde and the chemical entity



The above building block is disclosed in WO03078445A2, the content of which being incorporated herein in the entirety by reference.

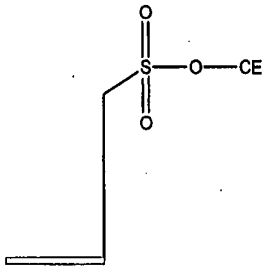
Building blocks transferring a chemical entity to a recipient reactive group forming a C-C bond

The below building block is able to transfer the chemical entity to a recipient group thereby forming a single bond between the receiving moiety, e.g. a scaffold, and the chemical entity.



The above building block is disclosed in WO03078445A2, the content of which being incorporated herein in the entirety by reference.

Another building block capable of transferring a chemical entity to a receiving reactive group forming a single bond is



The receiving group may be a nucleophile, such as a group comprising a hetero atom, thereby forming a single bond between the chemical entity and the hetero atom, or the receiving group may be an electronegative carbon atom, thereby forming a C-C bond between the chemical entity and the scaffold. The above building block is disclosed in WO03078446A2, the content of which is incorporated herein by reference.

The chemical entity attached to any of the above building blocks may be a selected from a large arsenal of chemical structures. Examples of chemical entities are H or entities selected among the group consisting of a C_1-C_6 alkyl, C_2-C_6 alkenyl, C_2-C_6 alkynyl, C_4-C_8 alkadienyl, C_3-C_7 cycloalkyl, C_3-C_7 cycloheteroalkyl, aryl, and heteroaryl, said group being substituted with 0-3 R^4, 0-3 R^5 and 0-3 R^9 or C_1-C_3 alkylene-$NR^4{}_2$, C_1-C_3 alkylene-$NR^4C(O)R^8$, C_1-C_3 alkylene-$NR^4C(O)OR^8$, C_1-C_2 al-

kylene-O-NR^4_2, C_1-C_2 alkylene-O-$NR^4C(O)R^8$, C_1-C_2 alkylene-O-$NR^4C(O)OR^8$ substituted with 0-3 R^9.

where R^4 is H or selected independently among the group consisting of C_1-C_6 alkyl, C_2-C_6 alkenyl, C_2-C_6 alkynyl, C_3-C_7 cycloalkyl, C_3-C_7 cycloheteroalkyl, aryl, heteroaryl, said group being substituted with 0-3 R^9 and

R^5 is selected independently from $-N_3$, -CNO, $-C(NOH)NH_2$, -NHOH, $-NHNHR^6$, $-C(O)R^6$, $-SnR^6_3$, $-B(OR^6)_2$, $-P(O)(OR^6)_2$ or the group consisting of C_2-C_6 alkenyl, C_2-C_6 alkynyl, C_4-C_8 alkadienyl said group being substituted with 0-2 R^7,

where R^6 is selected independently from H, C_1-C_8 alkyl, C_3-C_7 cycloalkyl, aryl or C_1-C_6 alkylene-aryl substituted with 0-5 halogen atoms selected from -F, -Cl, -Br, and –I; and

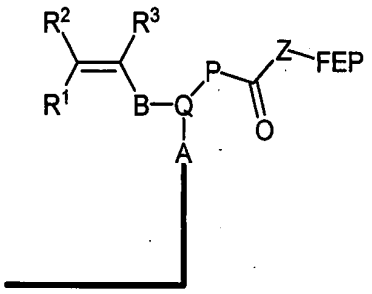
R^7 is independently selected from $-NO_2$, $-COOR^6$, $-COR^6$, -CN, $-OSiR^6_3$, $-OR^6$ and $-NR^6_2$.

R^8 is H, C_1-C_6 alkyl, C_2-C_6 alkenyl, C_2-C_6 alkynyl, C_3-C_7 cycloalkyl, aryl or C_1-C_6 alkylene-aryl substituted with 0-3 substituents independently selected from –F, -Cl, $-NO_2$, $-R^3$, $-OR^3$, $-SiR^3_3$

R^9 is =O, -F, -Cl, -Br, -I, -CN, $-NO_2$, $-OR^8$, $-NR^8_2$, $-NR^8$-C(O)R^8, $-NR^8$-C(O)OR^8, $-SR^8$, $-S(O)R^6$, $-S(O)_2R^6$, $-COOR^6$, $-C(O)NR^8_2$ and $-S(O)_2NR^8_2$.

Cross-link cleavage building blocks

It may be advantageous to split the transfer of a chemical entity to a recipient reactive group into two separate steps, namely a cross-linking step and a cleavage step because each step can be optimized. A suitable building block for this two-step process is illustrated below:



Initially, a reactive group appearing on the chemical entity precursor (abbreviated FEP) reacts with a recipient reactive group, e.g. a reactive group appearing on a scaffold, thereby forming a cross-link. Subsequently, a cleavage is performed, usually by adding an aqueous oxidising agent such as I_2, Br_2, Cl_2, H^+, or a Lewis acid. The cleavage results in a transfer of the group HZ-FEP- to the recipient moiety, such as a scaffold.

In the above formula

Z is O, S, NR^4

Q is N, CR^1

P is a valence bond, O, S, NR^4, or a group C_{5-7}arylene, C_{1-6}alkylene, C_{1-6}O-alkylene, C_{1-6}S-alkylene, NR^1-alkylene, C_{1-6}alkylene-O, C_{1-6}alkylene-S option said group being substituted with 0-3 R^4, 0-3 R^5 and 0-3 R^9 or C_1-C_3 alkylene-NR^4_2, C_1-C_3 alkylene-$NR^4C(O)R^8$, C_1-C_3 alkylene-$NR^4C(O)OR^8$, C_1-C_2 alkylene-O-NR^4_2, C_1-C_2 alkylene-O-$NR^4C(O)R^8$, C_1-C_2 alkylene-O-$NR^4C(O)OR^8$ substituted with 0-3 R^9,

B is a group comprising D-E-F, in which

D is a valence bond or a group C_{1-6}alkylene, C_{1-6}alkenylene, C_{1-6}alkynylene, C_{5-7}arylene, or C_{5-7}heteroarylene, said group optionally being substituted with 1 to 4 group R^{11},

E is, when present, a valence bond, O, S, NR^4, or a group C_{1-6}alkylene, C_{1-6}alkenylene, C_{1-6}alkynylene, C_{5-7}arylene, or C_{5-7}heteroarylene, said group optionally being substituted with 1 to 4 group R^{11},

F is, when present, a valence bond, O, S, or NR^4,

A is a spacing group distancing the chemical structure from the complementing element, which may be a nucleic acid,

R^1, R^2, and R^3 are independent of each other selected among the group consisting of H, C_1-C_6 alkyl, C_2-C_6 alkenyl, C_2-C_6 alkynyl, C_4-C_8 alkadienyl, C_3-C_7 cycloalkyl, C_3-C_7 cycloheteroalkyl, aryl, and heteroaryl, said group being substituted with 0-3 R^4, 0-3 R^5 and 0-3 R^9 or C_1-C_3 alkylene-NR^4_2, C_1-C_3 alkylene-$NR^4C(O)R^8$, C_1-C_3 alkylene-$NR^4C(O)OR^8$, C_1-C_2 alkylene-O-NR^4_2, C_1-C_2 alkylene-O-$NR^4C(O)R^8$, C_1-C_2 alkylene-O-$NR^4C(O)OR^8$ substituted with 0-3 R^9,

FEP is a group selected among the group consisting of H, C_1-C_6 alkyl, C_2-C_6 alkenyl, C_2-C_6 alkynyl, C_4-C_8 alkadienyl, C_3-C_7 cycloalkyl, C_3-C_7 cycloheteroalkyl, aryl, and heteroaryl, said group being substituted with 0-3 R^4, 0-3 R^5 and 0-3 R^9

or C_1-C_3 alkylene-NR^4_2, C_1-C_3 alkylene-$NR^4C(O)R^8$, C_1-C_3 alkylene-$NR^4C(O)OR^8$, C_1-C_2 alkylene-O-NR^4_2, C_1-C_2 alkylene-O-$NR^4C(O)R^8$, C_1-C_2 alkylene-O-$NR^4C(O)OR^8$ substituted with 0-3 R^9,

where R^4 is H or selected independently among the group consisting of C_1-C_6 alkyl, C_2-C_6 alkenyl, C_2-C_6 alkynyl, C_3-C_7 cycloalkyl, C_3-C_7 cycloheteroalkyl, aryl, heteroaryl, said group being substituted with 0-3 R^9 and

R^5 is selected independently from $-N_3$, -CNO, $-C(NOH)NH_2$, -NHOH, $-NHNHR^6$, $-C(O)R^6$, $-SnR^6_3$, $-B(OR^6)_2$, $-P(O)(OR^6)_2$ or the group consisting of C_2-C_6 alkenyl, C_2-C_6 alkynyl, C_4-C_6 alkadienyl said group being substituted with 0-2 R^7,

where R^6 is selected independently from H, C_1-C_6 alkyl, C_3.C_7 cycloalkyl, aryl or C_1-C_6 alkylene-aryl substituted with 0-5 halogen atoms selected from -F, -Cl, -Br, and –I; and R^7 is independently selected from $-NO_2$, $-COOR^8$, $-COR^6$, -CN, $-OSiR^6_3$, $-OR^6$ and $-NR^6_2$.

R^8 is H, C_1-C_6 alkyl, C_2-C_6 alkenyl, C_2-C_6 alkynyl, C_3-C_7 cycloalkyl, aryl or C_1-C_6 alkylene-aryl substituted with 0-3 substituents independently selected from –F, -Cl, $-NO_2$, $-R^3$, $-OR^3$, $-SiR^3_3$

R^9 is =O, -F, -Cl, -Br, -I, -CN, $-NO_2$, $-OR^6$, $-NR^6_2$, $-NR^6$-$C(O)R^8$, $-NR^6$-$C(O)OR^8$, $-SR^6$, $-S(O)R^6$, $-S(O)_2R^6$, $-COOR^6$, $-C(O)NR^6_2$ and $-S(O)_2NR^6_2$.

In a preferred embodiment Z is O or S, P is a valence bond, Q is CH, B is CH_2, and R^1, R^2, and R^3 is H. The bond between the carbonyl group and Z is cleavable with aqueous I_2.

Partitioning conditions

The partition step may be referred to as a selection or a screen, as appropriate, and includes the screening of the library for encoded molecules having predetermined desirable characteristics. Predetermined desirable characteristics can include binding to a target, catalytically changing the target, chemically reacting with a target in a manner which alters/modifies the target or the functional activity of the target, and covalently attaching to the target as in a suicide inhibitor.

The target can be any compound of interest. E.g. the target can be a protein, peptide, carbohydrate, polysaccharide, glycoprotein, hormone, receptor, antigen, antibody, virus, substrate, metabolite, transition state analogue, cofactor, inhibitor, drug, dye, nutrient, growth factor, cell, tissue, etc. without limitation. Particularly preferred

targets include, but are not limited to, angiotensin converting enzyme, renin, cyclooxygenase, 5-lipoxygenase, IIL- 1 0 converting enzyme, cytokine receptors, PDGF receptor, type II Inosine monophosphate dehydrogenase, β-lactamases, integrin, and fungal cytochrome P-450. Targets can include, but are not limited to, bradykinin, neutrophil elastase, the HIV proteins, including *tat, rev, gag, Int,* RT, nucleocapsid etc., VEGF, bFGF, TGFβ, KGF, PDGF, thrombin, theophylline, caffeine, substance P, IgE, sPLA2, red blood cells, glioblastomas, fibrin clots, PBMCs, hCG, lectins, selectins, cytokines, ICP4, complement proteins, etc.

Encoded molecules having predetermined desirable characteristics can be partitioned away from the rest of the library while still attached to the identifier nucleic acid sequence by various methods known to one of ordinary skill in the art. In one embodiment of the invention the desirable products are partitioned away from the entire library without chemical degradation of the attached nucleic acid identifier such that the identifiers are amplifiable. The identifiers may then be amplified, either still attached to the desirable encoded molecule or after separation from the desirable encoded molecule.

In a preferred embodiment, the desirable encoded molecule acts on the target without any interaction between the nucleic acid attached to the desirable encoded molecule and the target. In one embodiment, the bound complex-target aggregate can be partitioned from unbound complexes by a number of methods. The methods include nitrocellulose filter binding, column chromatography, filtration, affinity chromatography, centrifugation, and other well known methods.

Briefly, the library of complexes is subjected to the partitioning step, which may include contact between the library and a column onto which the target is immobilised. Identifier nucleic acids associated with undesirable encoded molecules, i.e. encoded molecules not bound to the target under the stringency conditions used, will pass through the column. Additional undesirable encoded molecules (e.g. encoded molecules which cross-react with other targets) may be removed by counter-selection methods. Desirable complexes are bound to the column and can be eluted by changing the conditions of the column (e.g., salt, pH, surfactant, etc.) or the identifier.

Additionally, encoded molecules which react with a target can be separated from those products that do not react with the target. In one example, a chemical compound which covalently attaches to the target (such as a suicide inhibitor) can be washed under very stringent conditions. The resulting complex can then be treated with proteinase, DNAse or other suitable reagents to cleave a linker and liberate the nucleic acids which are associated with the desirable chemical compound. The liberated nucleic acids can be amplified.

In another example, the predetermined characteristic of the desirable product is the ability of the product to transfer a chemical group (such as acyl transfer) to the target and thereby inactivate the target. One could have a product library where all of the products have a thioester chemical group. Upon contact with the target, the desirable products will transfer the chemical group to the target concomitantly changing the desirable product from a thioester to a thiol. Therefore, a partitioning method which would identify products that are now thiols (rather than thioesters) will enable the selection of the desirable products and amplification of the nucleic acid associated therewith.

There are other partitioning and screening processes, which are compatible with this invention that are known to one of ordinary skill in the art. In one embodiment, the products can be fractionated by a number of common methods and then each fraction is then assayed for activity. The fractionization methods can include size, pH, hydrophobicity, etc.

Inherent in the present method is the selection of encoded molecules on the basis of a desired function; this can be extended to the selection of molecules with a desired function and specificity. Specificity can be required during the selection process by first extracting identifier nucleic acid sequences of chemical compounds which are capable of interacting with a non-desired "target" (negative selection, or counter-selection), followed by positive selection with the desired target. As an example, inhibitors of fungal cytochrome P-450 are known to cross-react to some extent with mammalian cytochrome P-450 (resulting in serious side effects). Highly specific inhibitors of the fungal cytochrome could be selected from a library by first removing those products capable of interacting with the mammalian cytochrome, followed by

retention of the remaining products which are capable of interacting with the fungal cytochrome.

Brief Description of the Figures

Fig. 1 illustrates the overall process of building block evolution.

Fig. 2 shows the distribution of codon in different positions in an output from a selection.

Fig. 3 shows the difference between identifier driven and building block driven evolution.

Fig. 4 shows a method for reducing the library diversity through codon analysis.

Fig. 5 discloses two embodiments of using a Taqman probe (5' nuclease probe) in the measurement of the presence or absence of a certain codon.

Fig. 6 shows a standard curve referred to in example 4.

Fig. 7 shows a result of example 4.

Fig. 8 discloses a result of example 4.

Fig. 9 discloses a scheme relating to combined structural information and codon abundances in library design.

Fig. 10 discloses a relationship between codon analysis and structural information.

Fig. 11 shows the detection of single codons of identifiers.

Fig. 12 shows the detection of codon pairs of identifiers.

Fig. 13 shows the detection of codon pairs at specific codon positions.

Fig. 14 shows the detection of single codons of identifiers after the separation of the individual codons.

Fig. 15 discloses a method for selecting from a library, complexes capable of binding to a target molecule.

Fig. 16 discloses a method for enriching specific nucleic acid fragments and the utility of these fragments for the generation of a new library.

Fig. 17 discloses a method for reducing the diversity of a library of complexes.

Detailed description of the figures

Fig. 1A Shows the principle steps in BB evolution. An initial library of desired size is produced. This initial library is subjected to a selection process where encoded molecules that associate with a target of interest are enriched. The encoding identi-

fier oligonucleotide is preferably amplified and the used in the codon analysis step. This step monitors the relative abundance of each codon in the selected library. The information obtained in this analysis is used to design a new enriched library, which contains the preferable chemical entities and their corresponding codons. This new library is then subjected to a new selection process to select for binders. This diversity reduction cycle can be repeated until the desirable result is obtained and the binders have been obtained.

Fig. 1B shows how the diversity of a library (n^4) is reduced by reducing the number of chemical entities (n) in the library. Thus, by removing chemical entities not involved in the encoded molecules partitioned, a reduction in library diversity can be obtained to allow the identification of binders.

The identifier oligonucleotide that encodes for the display molecule is composed of codons and associated with the encoded molecule, as shown in Fig. 2. These codons possess information about the chemical entities in the encoded molecule. Each of these codon positions can be analysed for the precise sequence, which will reflect which chemical entities that have been enrich for in the selection process. The relative amount can also be obtained by comparing the signal in the measuring procedure (e.g. QPCR and array analysis). Each codon position will have its own fingerprint on which chemical entities that the selected display molecules possess. These fingerprints in each position can subsequently be used to put together a new more focused library with a lower and more enrich diversity that can be subjected to another round of selection. This can then repeated until the preferable encoded molecules have been obtained.

Fig. 3 illustrates the main difference between identifier and chemical entity (CE) evolution. In both cases the initial selection starts on a library with certain diversity. After the first round of selection the encoding identifiers are amplified where the distribution is maintained. This distribution is then transferred to the next generation which is used in a new selection. Thus, the strongest binders that were enriched in the first round of selection will be present at a relatively higher concentration compared to the weaker binders and the background. In the CE-driven evolution the codon analysis is used to design a new library. In this example, the new library is constructed to contain all the chemical entities that were identified as a positive signal in

the analysis. In other words, all the chemical entities that were not detected through the codon analysis were excluded in the new library. The new library is designed to have an equal amount of each selected chemical entity, which will generate all the possible display molecules at the same concentration. This will allow all binders to compete at the same concentration and potentially retain a more diverse set of binders in each round of selection. This is especially important for small molecules here not only the affinity is of interest.

Figure 4. This illustrates the process where the diversity is reduced through the codon analysis. An initial library of 10^{10} (e.g. 317*317*317*317) library members is subjected to a selection. The enrich identifier oligonucleotides are amplified and used in the codon analysis. The codon analysis result is used to design a new 10^7 (e.g. 57*57*57*57) library where the enriched chemical entities are included. This new library is the again subjected to a selection process. The identifier oligonucleotides are amplified and used for codon analysis. This new codon analysis results is again used to design a new 10^4 (e.g. 10*10*10*10) library where the enriched chemical entities are included. Finally a last selection step is performed in this reduced diversity library to identify the binders.

A preferred embodiment of the invention utilizing a universal Taqman probe is shown in Fig.5. Four codons are shown (P1 through P4; bold pattern) along with flanking regions (light pattern). A universal Taqman probe anneals to a region adjacent to the codon region, but within the amplicon defined by the universal PCR primers Pr.1 and Pr. 2. These primers could be the same as used for amplification of the identifier oligonucleotides encoding binders after an enrichment process on a specific target. However, are minimal length identifiers preferred during the encoding process, the region involved in Taqman probe annealing could be appended to the library identifier oligonucleotides by e.g. overlap PCR, ligation, or by employing a long downstream PCR primer containing the necessary sequences. The added length corresponding to the region necessary for annealing of the Taqman probe would be form 20 to 40 nts depending on the type of TaqMan probe and T_A of the PCR primers. The Q-PCR reactions are preferably performed in a 96- or 384-well format on a real-time PCR thermocycling machine.

Panel A shows the detection of abundance of a specific codon sequence in position one. Similar primers are prepared for all codon sequences. For each codon sequence utilized to encode a specific BB in the library a Q-PCR reaction is performed with a primer oligonucleotide complementary to the codon sequence in question. A downstream universal reverse primer Pr. 2 is provided after the Taqman probe to provide for an exponential amplification of the PCR amplicon. The setup is most suited for cases where the codon constitutes a length corresponding to a length suitable for a PCR primer.

Panel B shows the detection of abundance of a specific codon sequence in a specific codon position using a primer, which is complementing a codon and a framing sequence. Similar primers are used for all the codons and framing sequences. For each codon sequence utilized to encode a specific BB at a specific codon position in the library a Q-PCR reaction is performed with an oligo complementary to the codon sequence in question as well as a short region up- or downstream of the codon region which ensures extension of the primer in a PCR reaction only when annealed to the codon sequence in that specific codon position. The number of specific primers and Q-PCR reactions needed to cover all codon sequences in all possible codon positions equals the number of codon sequences times the number of codon positions. Thus, monitoring the abundance of 96 different codon sequences in 4 different positions can be performed in a single run on four 96 wells micro titre plates (as shown in Panel B) or a single 384 well plate on a suitable instrument. This architecture allows for the decoding of a $8{,}5 \times 10^7$ library of different encoded molecules.

Quantification is performed relative to the amount of full-length PCR product obtained in a parallel control reaction on the same input material performed with the two external PCR primers Pr.1 + Pr. 2. Theoretically, a similar rate of accumulation of this control amplicon compared to the accumulation of a product utilizing a single codon + sequence specific primer would indicate a 100% dominance of this particular sequence in the position in question.

Although the setups shown in Panel A and B employ a Taqman probe strategy, other detection systems (SYBR green, Molecular Beacons etc.) could be utilized. In theory, multiplex reactions employing up to 4 different fluorofors in the same reaction could increase throughput correspondingly.

An example of how a deconvolution process of a library of encoded molecules occurs is described in the following. Imagine that at the end of a selection scheme a pool of 3 ligand families (and the corresponding coding identifiers) are dominating the population and present at approx. the same concentration. Three different chemical entities are present in the first position of the encoded compounds, and each of these chemical entities are present in combination with one unique chemical entity out of 3 different chemical entities in position P2. Only one chemical entity in position 3 gives rise to active binders, whereas any of a 20% subset of chemical entities (*e.g.* determined by charge, size or other characteristics) is present in position 4. The outcome of the initial codon profile analysis would be: 3 codon sequences are equally dominating in position P1, 3 other codon sequences in position P2, 1 unique codon sequence is dominant in P3 whereas somewhat similarly increased levels of 20% of the codon sequences (background levels of the remaining 80% sequences) are seen in P4. In such cases it could be relevant to use an iterative Q-PCR ("IQ-PCR") strategy to perform a further deconvolution of a library after selection. Again with reference to the example above, by taking the PCR products from the 3 individual wells that contained primers giving the high yields in position P1, diluting the product appropriately and performing a second round of Q-PCR on each of these identifier oligonucleotides separately, it would be possible to deduce which codon sequence(s) is preferred in P2 when a given codon sequence is present in P1.

Fig. 9. This figure illustrates the possibility to combine structural information about the chemical entities and the relative abundance when designing a new more focused library. The structural information about the chemical entities can be used at least in two ways. First the similarities between the chemical entities in each position can be used to choose chemical entities to a new library. Secondly, the combination of the selected chemical entities can be analyzed to investigate possible pattern that generate potential ligands. This is especially useful if the binding site or the structure of a known ligand is known. Any type of structural analysis tool can be used that generate information about the structure of separate chemical entities or combination of chemical entities (the potential binders). By combining these three analysis approaches a more focused library can be generated that potentially will contain more specific binders compare to background binders. This new focused library can

be used in another round of selection to reduce the diversity. This procedure can be repeated until the desired binders have been identified.

Figure 10. This figure shows how the combination of codon analysis and structural information can generate valuable information. This invention allows the performance of structure activity relationship analysis (SAR) where the relative abundance in the codon analysis will represent the activity parameter (e.g. IC_{50} values) in the SAR measurements. Pharmacophore models can be generated, focused libraries can be designed, certain follow up chemistry can be used and information in the hit to lead process can be used.

Fig. 11 shows an array detection system in which a single codon is detected. Initially a library of selected complexes (29), i.e. complexes comprised of the initial library, which display a certain property, is provided as disclosed above. The initial library of complexes is prepared from e.g. 100 codons and identifiers having 4 codons in sequence, which theoretical gives a library of 10^8 complexes. The selected complexes are subjected to amplification to amplify the identifiers of the selected complexes and the amplification products are added to an array (30). The array (30) comprises probes (32) complementary to each of the codons of the identifiers (31). At hybridisation conditions the PCR products of the identifiers are annealed to the cognate probes of the array and in a suitable scanner the spatial position of the annealed probes are detected to elucidate the codons (33) of the identifier. The quantity of each codon may be measured to find codons abundant in more than one identifier and/or codons leading to encoded molecules with high affinity. The information may be used for decoding of the encoded molecule of the complexes displaying the desired property or the information may be used for selection of building blocks, which is to be added in a next round of library formation.

Fig. 12 discloses an array detection system for establishing codons pairs, i.e. codons in the vicinity of each other. Initially (as shown in this example) a library of complexes is prepared from 100 different codons deposited on an identifier in a sequence of four, making the total amount of combinations possible 10^8. The initial library is subjected to a condition in order to select a sub-library (29) displaying a desired property. The identifiers of the sub-library are amplified by a PCR reaction and the reaction product is added under hybridisation conditions to an array (34).

The array is designed with probes (35) capable of detecting two codons at a time. To cover all possible combinations of a library based on 100 different codons 10^4 probes are needed, which is practically feasible with the current technology. The detection of the codons may be conducted quantitatively, i.e. the relative abundance of each of the codon pairs may be determined. The detection on the array may be used to reconstruct the selected identifiers (36) as three overlapping codon pair detections depict the entire identifier. In the event the same codon pair appears on more than one identifier, the information on the relative abundance of each codon pair maybe used to decipher the sequence of codons of the selected identifiers as it can be assumed that each codon pair of the same identifier appears in the same amounts in the PCR products added to the array.

Fig. 13 discloses an array for detecting codon pairs at specific codon positions. Initially, a library of complexes comprising identifiers with framing sequences is provided. The framing sequence is specific for each position of the codons on the identifier. Four times more probes on the microarray is needed per each codon if the position of the codons also should be detected in the analysis which is practically feasible with current technology. The position is detected due to the framing sequences next to each codon. The initial library is subjected to a selection process to isolate complexes (37) having a desired property. The selected complexes are amplified by a PCR reaction and the reaction products are added to an array (38). The array comprises probes capable of detecting codon pairs as wells as the framing sequences (40) between the codons. The framing sequence determines the position of the codon in the reaction history, i.e. it is possible to deduct which chemical entity that reacted at which point in time of the synthesis history of the encoded molecule, thus making it possible to reconstruct the structure of the encoded molecule.

The detection of the codon pairs may be conducted quantitatively, i.e. the relative abundance of each of the codon pairs may be determined. The detection on the array may be used to reconstruct the selected identifiers (41) as three overlapping codon pair detections depict the entire identifier. In the event the same codon pair appears on more than one identifier, the information on the relative abundance of each codon pair maybe used to decipher the sequence of codons of the selected identifiers as it can be assumed that each codon pair of the same identifier appears in the same amounts in the PCR products added to the array.

Fig. 14 shows an array detection system in which a single codon is detected. Initially a library of selected complexes (42), i.e. complexes comprised of the initial library which display a certain property, is provided as disclosed above. The initial library of complexes is prepared from e.g. 100 codons and identifiers having 4 codons in sequence, which theoretical gives a library of 10^8 complexes. The selected complexes are subjected to amplification to amplify the identifiers of the selected complexes and the amplification products are treated with suitable reagents to cut between the individual codons (43). The individual codon is the applied to the array. The array (44) comprises probes (45) complementary to each of the codons of the identifiers (46). At hybridisation conditions the PCR products of the identifiers are annealed to the cognate probes of the array and in a suitable scanner the spatial position of the annealed probes are detected to elucidate the codons (47) of the identifier. The quantity of each codon may be measured to find codons abundant in more than one identifier and/or codons leading to encoded molecules with high affinity. The information may be used for decoding of the encoded molecule of the complexes displaying the desired property or the information may be used for selection of building blocks, which is to be added in a next round of library formation.

Fig. 15 discloses a method for selection of a suitable complex in several steps. In a first step the library of complexes 1 is provided. Each member of the library comprises an encoded molecule 2 composed of four chemical entities which is attached to an identifier oligonucleotide 3, which comprises four codons. The initial library shown comprises three complexes. In a second step the library of complexes is incubated with immobilized target molecules 4. The encoded molecule having an affinity towards the target molecule is bound to the immobilized target whereas encoded molecules not having affinity towards the target under the conditions used remains in the liquid media. The complexes remaining in the liquid media are discarded by a washing process, while the bound complexes remain attached to the immobilized target molecules. The washing process is usually conducted using mild stringency conditions in the initial rounds of selection. In later stage selections the working stringency conditions are usually increased to allow only high affinity binders to remain attached to the target. Subsequent to the washing step the complexes having affinity towards the target molecule are recovered. The recovery process usually requires high stringency conditions to detach the encoded molecule from the

immobilized the target. The selected sub-library resulting from the elution is subjected to an amplification process. The amplification of the identifier nucleic acid sequence of the selected complexes is usually performed using the PCR method. Preferably, a modification of the PCR method is followed such that a biotin molecule is attached to one of the primers to obtain a handle for subsequent immobilization. The result of the amplification step is multiple copies of the identifier nucleic acid sequences, which codes for the encoded molecules which have survived the selection step.

Fig. 16 discloses an enrichment process of building blocks. The building blocks can be used for generation of a new library. Initially, identifier nucleic acid sequences are immobilized on solid support. In one aspect of the invention the identifier nucleic acid sequences are the product of the selection procedure described in Fig. 1. Each codon of the identifier nucleic acid sequence is identified with an uppercase letter, i.e. A, B, C, or D. The immobilized identifier acid sequences are contacted with the pool of building blocks under hybridisation conditions. Each of the building blocks are illustrated with an sequence complementary to a codon which may or may nor be present on the identifier nucleic acid sequence. The complementary sequences are indicated with a apostrophe, e.g. A', B', etc. The transferable chemical entity of a building block is illustrated with a lowercase letter. The conditions providing for hybridisation of the complementing sequences of the pool of building blocks to the immobilised identifier nucleic acid sequence are preferably such that cognate nucleic acid sequences are hybridised to each other while sequences not recognizing any immobilized sequence remain in aqueous media. The immobilized sequences of the identifier nucleic acid sequences are thus used as bait in catching building blocks with complementing sequences. Following the incubation step, non-binding building bocks are removed by washing, whereby the part of the pool of building blocks not being able to find a complementing sequence is discarded. The building blocks attached to the immobilized nucleic acid sequences are detached using dehybridisaton conditions. The diminished pool of building blocks may be used in a subsequent round for preparing a new library of complexes, in which the encoded molecule comprises a reaction product comprising additions from chemical entities attached to the enriched building blocks. Because the order of building blocks which have participated in the formation of the encoded molecules successful in the selection procedure, is not preserved by the method for enriching building blocks a

scrambling of the encoded molecules may be obtained in some of the methods described herein for obtaining a library of complexes. In some applications of the library it will be an advantage to have a scrambling of the building blocks because and increased diversity is obtained.

Fig. 17 discloses a method for reducing the diversity of the library of complexes resulting from the method described in Fig. 16. In some of the applications of the library the diversity induced by scrambling of the building blocks are not desired. In a first step the sequences complementary to the identifier acid sequences used in Fig. 16 are provided and immobilized on a suitable solid support. In one aspect of the invention the complementary sequence is obtained from the PCR product resulting from the method according to Fig. 15. Alternatively, the complementing sequence may be obtained by extending the identifier nucleic acid sequence using a suitable primer, optionally attached to a handle such as a biotin or dinitrophenol. In a second step the immobilized complementary sequence is incubated with the scrambled library under conditions, which provide for hybridisation between the complementary sequence and members of the library having affinity towards this sequence. Members of the library not having affinity to the complementary sequences remains in the media and is discarded, while members of the library being able to hybridise to the immobilized nucleic acid sequences is recovered. Occasionally, nucleic acids not perfectly matching with the complementary sequence immobilized on the solid support are caught. In one aspect of the invention the hybridisation products, prior to the recovery step, are treated with an enzyme capable of recognizing mismatching nucleotides and cleaving the double stranded helix in which they are situated. An example of an enzyme with this ability is T4 endonuclease VII. After the treatment with the enzyme, complexes displaying a hybridisation toward the immobilized sequence are eluted under dehybridisation conditions. Nucleotide sequences remaining from the cleavage by the enzyme will also be present in the new library, however, these sequences will not have any effect of a subsequent selection because no molecule is attached thereto.

Examples

Example 1. Enrichment of nucleic acid fragments

A codon was included in the oligonucleotide sequence shown below. The codon is underlined and the boldface sequences represent the "framing" regions next to each codon. These framing regions can be used for specifying the position of each codon.

Biotin-AATTCCGGAA**CATA**CTAGTCAAC**ATGA**-3' (SEQ ID NO:1)

This identifier oligonucleotide was immobilized on streptavidin beads using standard protocols, i.e. 600 pmol identifier oligonucleotide with 5'- dT biotin in 50 µl 100 mM Mes pH 6.0 was mix with 50 µl SA-magnetic beads (Roche). The mixture was washed 2-3 times with 100 mM MES pH 6.0 to remove non-bound identifier oligonucleotides. To reduce background binding, the oligos and beads was incubated at RT for 10 min on shaker, then incubated on ice for 10 min while rotating the tube. Finally, the sample was washed with 100 mM MES 4 times in 800 µl at 60°C.

In the case where a PCR product is immobilized, the complementing (non-sense) strand is removed using 10 mM NaOH. This will generate single-stranded DNA with the selected codons. The same procedure described in this example can be used for a collection of different identifier nucleic acid molecules that contain one or more codons. The codons in the identifier nucleic acid molecules can be the same or different determined from the enrichment performed on the initial library.

The immobilized identifier nucleic acid molecule was mixed with the pool of nucleic acid fragments shown below. This pool of fragments illustrates an original pool that was used for generating an initial library of complexes. Each fragment may possess in the 3'-end a specific chemical entity that is encoded by the codon sequence. These nucleic acid fragments contain a specific sequence in the codon region (underlined) while the framing region shown in boldface is identical among the fragments. Thus, the pool of fragments represents different codons in the same position of the identifier nucleic acid.

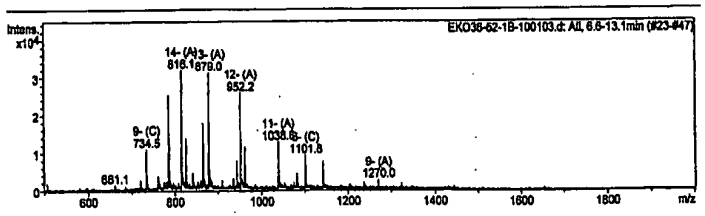
1. CGT GTG ATC GAA CTC GTG TG **GTAT**GATCAGTTG**TACT**-5' (SEQ ID NO:2)

2. CGT GTG ATC GAA CTC GTG TG **GTAT**CTAGTCGGT**TACT**-5' (SEQ ID NO:3)

3. CGT GTG ATC GAA CTC GTG TG **GTAT**TCGAGTGTT**TACT**-5' (SEQ ID NO:4)

4. CGT GTG ATC GAA CTC GTG TG **GTAT**AGCTCATGG**TACT**-5' (SEQ ID NO:5)

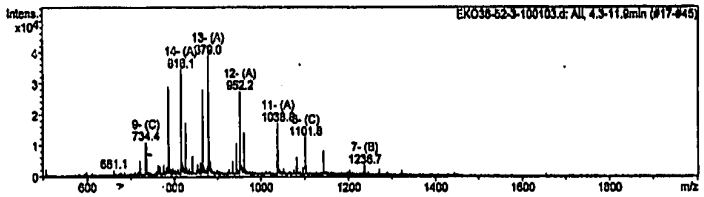
The nucleic acid fragments are mixed with the immobilized identifier nucleic acid using 600 pmol of each nucleic acid fragment mixed with the immobilized identifier nucleic acid molecules (100 mM MES pH 6.0, 150 mM NaCl)). The mixture was incubated at 25°C for 30 minutes in a shaker. The non-hybridized fragments were removed by 4 times washing in 800 µl 100 mM MES, 150 mM NaCl. This step should separate the complementing fragments (bound) encoding for the select chemical entities from the non-complementing fragments (non-bound) encoding for chemical entities that were not effective in the preceding selection process. The annealed fragments are eluted from the immobilized identifier nucleic acid molecules by re-suspending the beads in 25 µl 60°C H_2O and incubating for 2 min at 60°C. The enriched fragments were purified on a micro-spin gel filtration column (BiRad). The eluted fragments were prepared for mass spectroscopy (MS) analysis by mixing in half volume of ion exchanger resin and incubating minimum 2 h at 25°C on a shaker. After incubation the resin was removed by centrifugation and 15 µl of the supernatant was mixed with 7 µl of water, 2 µl of piperidine and imidazole (each 625 mM) and 24 µl acetonitrile. The sample was analysed using a Mass Spectroscopy instrument (Bruker Daltonics, Esquire 3000plus). The result for the MS analysis is shown below.



Component	Deconvoluted Mass	Molecule	Absolute Abundance	Relative Abundance
A	11438.39	[M - H]-	112674	96.98 ⇒ MATCH
B	8663.73	[M - H]-	61649	53.06
C	6616.63	[M - H]-	41245	35.50

The mass from the correct complementary fragment (number 1) is obtained in the MS analysis (11438.39, expected 11439 Da) No masses for the other fragments (number 2-4) could not be found in the MS spectra (expected masses; 11415, 11430, 11424 Da). This result shows that the right fragment is strongly enriched and other fragments with the wrong codon sequences are removed. The enrichment is possible even when the "spacing" region (boldface) is identical in each fragment.

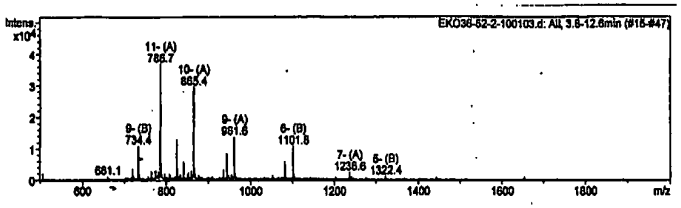
Two control experiments were also performed to validate the enrichment protocol. In the first experiment, the fragment with the correct codon sequence (number 1) was mixed with the immobilized identifier molecule as described above. The sample was washed end eluted also as described above and prepared for MS analysis. The result from the MS analysis is shown below.



Component	Deconvoluted Mass	Molecule	Absolute Abundance	Relative Abundance
A	11438.39	[M - H]-	127110	96.21 ⇒ MATCH
B	8663.61	[M - H]-	82800	62.67
C	6616.52	[M - H]-	51480	38.97

The result indicates that the fragment with the correct sequence (number 1) anneals to the immobilized identifier molecules and is eluted under the conditions used in this example. The expected mass (11439) correlate well with the experimental mass, 11438.39 Da.

In the other control experiment, a fragment with a wrong codon sequence (number 3) was allow to bind to the immobilized identifier molecule as described above. Again, the eluted sample was prepared and analysed with MS. The result is shown below.



Component	Deconvoluted Mass	Molecule	Absolute Abundance	Relative Abundance
A	8663.47	[M - H]-	89308	96.60
B	8616.52	[M - H]-	41963	45.39
C	8678.68	[M - H]-	13651	14.77

In this experiment, no mass was found that corresponded to the expected mass (11430) of the tested building block (number 1). Again, this shows that fragments with a anticodon sequence different from the enriched codons in the identifier nucleic acid molecules are not captured using this approach.

The enriched fragments obtained using this strategy may then be used to generate a new library of encoded molecules. This new library will contain encoded molecules composed of the enriched chemical entities. Thus, the library size have been reduced due to the removal of chemical entities not involved in binding encoded molecules, and enriched in chemical entities that are highly represented in the encoded molecules which binds to the target molecule.

Example 1 shows the possibility of enriching for specific building block molecules, i.e. nucleic acid fragments associated with transferable chemical entities. The same procedure can be used for a larger pool of building block than four as used herein. The codon design will determine the maximum number of building blocks that can be used. The sequence in the codon region should be large enough to allow discrimination in the annealing step. Various conditions can be used to increase the stringency in the annealing step. Parameters such as temperature, salt, pH, formamide concentration, time and other conditions could be used.

Example 2 (model): Multiple codon selection in a library.

This example describes the enrichment of building blocks using an identifier nucleic acid (identifier) molecule with multiple codons. These codons encode for a displayed molecule (DM) that are attached to the identifier molecule before the selection is

performed. The library size is determined both by the number of different chemical entities and the total number of chemical entities. The identifier molecule shown below contains three codons. The codons, which codes for the displayed molecule are indicated with underlines and the region separating (framing region) the codons in boldface. The size of the codons can be varied dependent in the diversity need in the library and the optimal setup for chemical entity enrichment. The framing region can also be varied dependent on the discrimination needed to distinguish the precise position of a codon in the identifier molecule. The framing region will also be important for the generation of the library. This can be understood when the encoding is accomplished by extension of the encoding region as disclosed in DK PA 2002 01955 and US 60/434,425, incorporated herein by reference. There need to be a perfect match in the 3'-end in order to get efficient extension with a polymerase or a ligase. The size of this spacing/framing region should be long enough to form a complementing region to allow extension with a polymerase or ligase. Preferably, the spacing region should be between 3 and 6 nucleotides. The codon region together with the spacing region will also be useful when codons are to be identified using a micro array setup. The identifier molecule with the right codon sequences will hybridize to the array and be detected.

The sequence below represents an enriched identifier molecule attached to the displayed molecule (DM). This identifier molecule has been enriched due to the fact that the DM binds to the target molecule in the selection process. In practice, more than one enriched identifier molecules will be obtained when using a library of displayed molecules attached to its identifier sequence.

DM-
GCACACTAGCTTGAGCACACTGACACATGGAGATCACATGCTTCGACAATGCAGGACTCCCGCAGCTTTACGATCCCGCAGGTAACC
GT

This identifier molecule is amplified with two primers (below) using a standard PCR reaction. For example, 500 nM of each primer, 2,5 units Taq polymerase, 0.2 mM of each NTP, in a PCR buffer (50 mM KCl, 10 mM Tris-Cl, 3 mM DTT, 1.5 mM $MgCl_2$, 0.1 mg/ml BSA). Run 25 cycles (94°C melt for 30 seconds, 55°C anneal for 45 seconds, 72°C extension for 60 seconds).

B-GCACACTAGCTTGAGCACACTGACA-3'
CGAAATGCTAGGGCGTCCATTGGCA-5'

This will amplify the identifier molecule from the selection process and add a biotin in the 5'-end of one of the strand (below). This amplified product is then immobilized on a solid support, streptavidin beads for example. This can be performed identical as describe in example 1.

When the identifier molecules have been immobilized and the excess has been removed by a washing step (as describe in example 1), the complementing non-sense stand is removed by incubating in 10 NaOH for about 2 min and washed with 100 mM Mes buffer, pH 6.0. This procedure will generate the strand shown below where the codon regions are exposed to allow hybridization with the complementing sequences.

```
B-
GCACACTAGCTTGAGCACACTGACACATGGAGATCACATGCTTCGACAATGCAGGACTCCCGCAGCTTTACGATCCCGCAGGTAACC
GT
```

The next step is to protect the complementing sequences outside the codons to prevent the binding of the building block to these sequences. This can be performed by adding "blocking" oligonucleotides that has a complementing sequence. This is shown below.

```
B-
GCACACTAGCTTGAGCACACTGACA                    CTTCGACAATGCAGGACTCCCGCAGCTTTACGATCCCCCAGGTAACC
GT
CGTGTGATCGAACTCGTGTGACTGT                                    CGAAATGCTAGGGCGTCCATT-
GGCA
```

Next, the pool of different building blocks is added and is allowed annealing to the codon region in the identifier region. The position of annealing is determined by the spacing region shown in boldface. The stringency is adjusted to only allow hybridization of the correct building block in the right position. This can be accomplished by mixing the right component together using various conditions. The condition can for example include the presence of salt, formamide and various buffers adjusted to suitable pH and temperature. Below is the correct building block that will anneal to the enriched identifier molecules. These building blocks is annealed and eluted as described in example 1.

```
CE-CGTGTGATCGAACTCGTGTGACTGTGTACCTCTAGTGTAC
```


The next pool of building blocks is blocked with an oligonucleotide that also protects the first codon. This is necessary to prevent binding of the building blocks in that codon.

```
B-
GCACACTAGCTTGAGCACACTGACACATGGAGATCAC                    AGGACTCCCGCAGCTTTACGATCCCGCAGGTAACC
GT
CGTGTGATCGAACTCGTGTGACTGTGTAIIIIIIIII                                CGAAATGCTAGGGCGTCCATT-
GGCA
```

Again, the library of building blocks is added to enrich for the selected codons. Below is the building block with the correct sequence. These building blocks is annealed and eluted as described in example 1.

```
CE--CGTGTGATCGAACTCGTGTGACTGTGTAIIIIIIIIITACGAAGCTGTTACG
```

Finally, the identifier molecule is protected with a blocking oligo that expose only the last codon.

```
B--
GCACACTAGCTTGAGCACACTGACACATGGAGATCACATGCTTCGACAA                    AGGACTCCCGCAGCTTTACGATCCCGCAGGTAACC
GT
CGTGTGATCGAACTCGTGTGACTGTGTAIIIIIIIIITACIIIIIIIII                    CGAAATGCTAGGGCGTCCATT-
GGCA
```

A new pool of building blocks is added and allowed hybridizing to the identifier molecule. These building blocks is annealed and eluted as described in example 1.

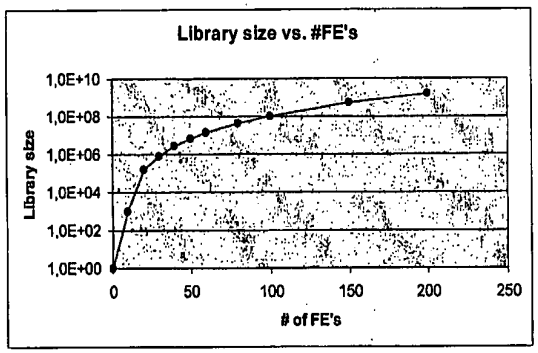
```
CE--CGTGTGATCGAACTCGTGTGACTGTGTAIIIIIIIIITACIIIIIIIIITACGTCCTGAGGGCGT
```

The enrichment of each library of building blocks are performed in separate tubes in order to keep the libraries of building block separated. The enrichment is performed with building blocks loaded with chemical entities (CE).

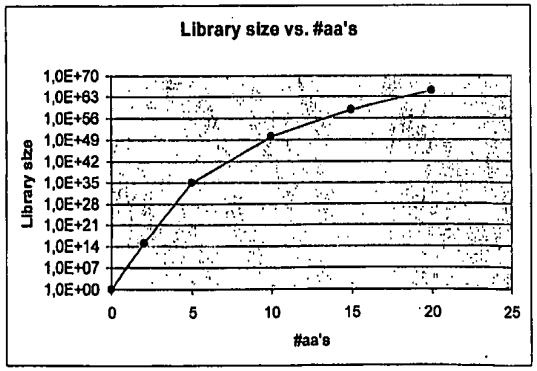
Example 3 – Template versus chemical entity evolution

The graph below illustrates the relationship between the number of chemical entities and the library size. The example below is calculated on that the final encoded molecules contains four chemical entities that is individually encoded by the corresponding building block (n^4, where n is the number of building blocks). The graph shows that the diversity decreases dramatically with the reduction of the total number of building blocks. If the number of different building can be reduced to about 20-

30 (library size of $16*10^3$ and $81*10^4$, respectively) in the selection process, then the library size for the final round of selection is low enough for identification of the binding molecules.



When the same analysis is performed on a protein another situation is obtained. The example shown below is on a very small protein (50 amino acids in length). The diversity is enormous when all amino acids are included in the library. The size of the library is also decreasing with the total number of amino acids, but not to the same extent as show above for a small molecule. Even when the different amino acids are reduced to 2, the library size is huge ($1.2\ 10^{15}$). This shows that amino acid enrichment is impossible on protein. This is even more pronounced for mid-size protein which contains about 300 amino acids.



Example 4 – Codon analysis

This example illustrates one possibility to perform codon analysis on a whole population of different identifier oligonucleotides. The analysis can also be performed using array where the probe oligonucleotides (complementary to the codons) are immobilized in discreet areas and the signal is monitored dependent on the amount of identifiers oligonucleotides are hybridised in each specific area. Codon analysis can also be performed using standard sequencing using a polymerase extension step.

In fig. 5, Four codons are shown (P1 through P4; bold pattern) along with flanking regions (light pattern). A universal Taqman probe anneals to a region adjacent to the codon region, but within the amplicon defined by the universal PCR primers Pr.1 and Pr. 2. These primers could be the same as used for amplification of the identifier oligonucleotides encoding binders after an enrichment process on a specific target. However, are minimal length identifiers preferred during the encoding process, the region involved in Taqman probe annealing could be appended to the library identifier oligonucleotides by e.g. overlap PCR, ligation, or by employing a long downstream PCR primer containing the necessary sequences. The added length corresponding to the region necessary for annealing of the Taqman probe would be form 20 to 40 nts depending on the type of TaqMan probe and T_A of the PCR primers. The Q-PCR reactions are preferably performed in a 96- or 384-well format on a real-time PCR thermocycling machine.

Fig. 5, panel A, shows the detection of abundance of a specific codon sequence in position one. Similar primers are prepared for all codon sequences. For each codon sequence utilized to encode a specific BB in the library a Q-PCR reaction is performed with a primer oligonucleotide complementary to the codon sequence in question. A downstream universal reverse primer Pr. 2 is provided after the Taqman probe to provide for an exponential amplification of the PCR amplicon. The setup is most suited for cases where the codon constitutes a length corresponding to a length suitable for a PCR primer.

Fig. 5, panel B shows the detection of abundance of a specific codon sequence in a specific codon position using a primer which is complementing a codon and a fram-

ing sequence. Similar primers are used for all the codons and framing sequences. For each codon sequence utilized to encode a specific BB at a specific codon position in the library a Q-PCR reaction is performed with an oligo complementary to the codon sequence in question as well as a short region up- or downstream of the codon region which ensures extension of the primer in a PCR reaction only when annealed to the codon sequence in that specific codon position. The number of specific primers and Q-PCR reactions needed to cover all codon sequences in all possible codon positions equals the number of codon sequences times the number of codon positions. Thus, monitoring the abundance of 96 different codon sequences in 4 different positions can be performed in a single run on four 96 wells micro titre plates (as shown in Fig. 5, panel B) or a single 384 well plate on a suitable instrument. This architecture allows for the decoding of a $8{,}5 \times 10^7$ library of different encoded molecules.

Quantification is performed relative to the amount of full-length PCR product obtained in a parallel control reaction on the same input material performed with the two external PCR primers Pr.1 + Pr. 2. Theoretically, a similar rate of accumulation of this control amplicon compared to the accumulation of a product utilizing a single codon + sequence specific primer would indicate a 100% dominance of this particular sequence in the position in question.

Although the setups shown in Fig. 5, panel A and B employ a Taqman probe strategy, other detection systems (SYBR green, Molecular Beacons etc.) could be utilized. In theory, multiplex reactions employing up to 4 different fluorofors in the same reaction could increase throughput correspondingly.

An example of how a deconvolution process of a library of encoded molecules occurs is described in the following. Imagine that at the end of a selection scheme a pool of 3 ligand families (and the corresponding coding identifiers) are dominating the population and present at approx. the same concentration. Three different chemical entities are present in the first position of the encoded compounds, and each of these chemical entities are present in combination with one unique chemical entity out of 3 different chemical entities in position P2. Only one chemical entity in position 3 gives rise to active binders, whereas any of a 20% subset of chemical entities (*e.g.* determined by charge, size or other characteristica) are present in position 4. The outcome of the initial codon profile analysis would be: 3 codon sequences are equally dominating in position P1, 3 other codon sequences in position P2, 1 unique codon sequence is dominant in P3 whereas somewhat similarly increased levels of 20% of the codon sequences (background levels of the remaining 80% sequences) are seen in P4. In such cases it could be relevant to use an iterative Q-PCR ("IQ-PCR") strategy to perform a further deconvolution of a library after selection. Again with reference to the example above, by taking the PCR products from the 3 individual wells that contained primers giving the high yields in position P1, diluting the product appropriately and performing a second round of Q-PCR on each of these identifier oligonucleotides separately, it would be possible to deduce which codon sequence(s) is preferred in P2 when a given codon sequence is present in P1.

Identifiers used for Q-PCR quantification

Taqman MGB probe binding region: *=AATTCCAGCTTCTAGGAAGAC

P1 P2 P3 P4

5'-CAGCTTGGACACCACGTCATACTAGCTGCTAGAGATGTGGTGATATTAGTGTGTGACGATGGTACGCACAAGTACGAACGTGCATCAGAGAGGACGAGCAGGACCTGGAACCTGGTGC*TTCCTCCACCACGTCTCTGAC-3'

GGAAGAAGACAGAAGACCTG
TCAGGAGTCGAGAACTGAAG
TGTGTACGTCAACACGTCAG
TGTGGAACTACCATCCAAGG
CCATCCAACATCGTTGGAAG
AACCTGTCCTGTGAGATCTG
TCACGAAGCTGGATGATGAG
TAGCATCGATCGAACGTAGG
TCGAAGCTACTGTCGAGATG

CTCGACCACTGCAGGTGGAGCTCC
CGTGCTTCCTCTGCTGCACCACCG
CCTGGTGTCGAGGTGAGCAGCAGC
CTCGACGAGGTCCATCCTGGTCGC
CGTGAGGAGCAGGTCCTCCTGTCG
CCTGACACTGGTCGTGGTCGAGGC
CCATCTCGACGACCTGCTCCTGGG
CCACGAGGTCTCCACTGGTCCAGG
CCACTGAGCTGCTCCTCCAGGTGG

Oligos for identifier synthesis:

FPv2: CAGCTTGGACACCACGTCATAC

RPv2: GTCAGAGACGTGGTGGAGGAA

Temp1-1: CAGCTTGGACACCACGTCATACTAGCTGCTAGAGATGTGGTGATATTAGTGTGTGACGAT
Temp1-2: CAGCTTGGACACCACGTCATACGGAAGAAGACAGAAGACCTGATATTAGTGTGTGACGAT
Temp1-3: CAGCTTGGACACCACGTCATACTCAGGAGTCGAGAACTGAAGATATTAGTGTGTGACGAT
Temp1-4: CAGCTTGGACACCACGTCATACTGTGTACGTCAACACGTCAGATATTAGTGTGTGACGAT
Temp1-5: CAGCTTGGACACCACGTCATACTGTGGAACTACCATCCAAGGATATTAGTGTGTGACGAT
Temp1-6: CAGCTTGGACACCACGTCATACCCATCCAACATCGTTGGAAGATATTAGTGTGTGACGAT
Temp1-7: CAGCTTGGACACCACGTCATACAACCTGTCCTGTGAGATCTGATATTAGTGTGTGACGAT
Temp1-8: CAGCTTGGACACCACGTCATACTCACGAAGCTGGATGATGAGATATTAGTGTGTGACGAT
Temp1-9: CAGCTTGGACACCACGTCATACTAGCATCGATCGAACGTAGGATATTAGTGTGTGACGAT
Temp1-10: CAGCTTGGACACCACGTCATACTCGAAGCTACTGTCGAGATGATATTAGTGTGTGACGAT

Temp2: GTCCTCTCTGATGCACGTTCGTACTTGTGCGTACCATCGTCACACACTAATATC

Temp3-1: GAACGTGCATCAGAGAGGACGAGCAGGACCTGGAACCTGGTGCAATTCCAGCTTCTAGGAAGACT
Temp3-2: GAACGTGCATCAGAGAGGACTCGACCACTGCAGGTGGAGCTCCAATTCCAGCTTCTAGGAAGACT
Temp3-3: GAACGTGCATCAGAGAGGACGTGCTTCCTCTGCTGCACCACCGAATTCCAGCTTCTAGGAAGACT
Temp3-4: GAACGTGCATCAGAGAGGACCTGGTGTCGAGGTGAGCAGCAGCAATTCCAGCTTCTAGGAAGACT
Temp3-5: GAACGTGCATCAGAGAGGACTCGACGAGGTCCATCCTGGTCGCAATTCCAGCTTCTAGGAAGACT
Temp3-6: GAACGTGCATCAGAGAGGACGTGAGGAGCAGGTCCTCCTGTCGAATTCCAGCTTCTAGGAAGACT
Temp3-7: GAACGTGCATCAGAGAGGACCTGACACTGGTCGTGGTCGAGGCAATTCCAGCTTCTAGGAAGACT
Temp3-8: GAACGTGCATCAGAGAGGACCATCTCGACGACCTGCTCCTGGGAATTCCAGCTTCTAGGAAGACT
Temp3-9: GAACGTGCATCAGAGAGGACCACGAGGTCTCCACTGGTCCAGGAATTCCAGCTTCTAGGAAGACT
Temp3-10: GAACGTGCATCAGAGAGGACCACTGAGCTGCTCCTCCAGGTGGAATTCCAGCTTCTAGGAAGACT

Temp4: GTCAGAGACGTGGTGGAGGAAGTCTTCCTAGAAGCTGGAATT

Synthesis of identifier oligonucleotides:

The 10 identifier oligonucleotides were assembled in 10 seperate 50 µl PCR reactions each containing 0.05 pmol of the oligos Q-Temp1-X, Q-Temp2, Q-Temp3-X and Q-Temp4 (x=1 through 10) and 25 pmol of the external primers FPv2 and RPv2 with TA= 53°C. The 160 bp products were gel-purified using QIAquick Gel Extraction Kit from QIAGEN (Cat. No. 28706) and quantified on spectrophotometer. As a control, 20 ng of each of the identifiers (as estimated from these measurements) were loaded on an agarose gel.

Preparation of samples for Q-PCR:

Sample A: Generated by mixing 20 ng from each identifier oligonucleotide prep. Volume was adjusted to 50 µl. Concentration: 4 ng/µl = 38.46 fmol/µl (160bp x 650 Da/bp =1.04x105 g/mol. 1 ng= 9.615 fmol). Diluted to 10^7 copies/5µl (0.00332 fmol/µl).

Sample B: 20 ng/20µl stocks of each identifier were prepared. The sample was mixed as follows:

5µl undil. Identifier #10
5µl 2x dil. Identifier #9
5µl 4x dil. Identifier #8
5µl 8x dil. Identifier #7
5µl 16x dil. Identifier #6
5µl 32x dil. Identifier #5
5µl 64x dil. Identifier #4
5µl 128x dil. Identifier #3
5µl 256x dil. Identifier #2
5µl 512x dil. Identifier #1

Concentration: 10ng/50µl= 0.20 ng/µl = 1.923 fmol/µl. Diluted 579.2-fold to 10^7 copies/5µl (0.00332 fmol/µl).

Standard curve: The samples for the standard curve was prepared by diluting Sample A 116.55-fold to 10^9 copies/5 µl (0.33 fmol//µl) and subsequently performing a 10-fold serial dilution of this sample. 5 µl was used for each PCR reaction. The standard curve is shown in Fig. 2.

Q-PCR reactions

For 5 ml premix (for one 96-well plate):

2.5 ml Taqman Universal PCR Master Mix (Applied Biosystems; includes Taq polymerase, dNTPs and optimized Taq pol. buffer)
450 µl RPv2 (10 pmol/ul)
25 µl Taqman probe (6-FAM-TCCAGCTTCTAGGAAGAC-MGBNFQ; 50 µM; Applied Biosystems)
1075 µl H2O

40.5 µl premix was aliquoted into each well and 4.5 µl of relevant upstream PCR primer (FPv2 (for standard curve) or one of the codon specific primers listed below; 10 pmol/µl) and 5 µl sample (H2O in wells for negative controls) was added. The codon-specific PCR primers were: (Tm calculations shown are from Vector NTI; matched to Tm for RPv2 (67.7°C))

P1-1:	GTCATACTAGCTGCTAGAGATGTGGTGATA	66.8°C
P1-2:	CATACGGAAGAAGACAGAAGACCTGATA	67.8°C
P1-3:	TCATACTCAGGAGTCGAGAACTGAAGATA	67.6°C
P1-4:	CATACTGTGTACGTCAACACGTCAGATA	67.4°C
P1-5:	CATACTGTGGAACTACCATCCAAGGATA	68.0°C
P1-6:	CCATCCAACATCGTTGGAAGAT	67.8°C
P1-7:	CATACAACCTGTCCTGTGAGATCTGATA	67.7°C
P1-8:	ATACTCACGAAGCTGGATGATGAGATA	67.3°C
P1-9:	CATACTAGCATCGATCGAACGTAGGATA	68.1°C
P1-10:	TCATACTCGAAGCTACTGTCGAGATGATA	68.2°C
P2-1:	ATATTAGTGTGTGACGATGGTACGCA	67.8°C
P3-1:	ACAAGTACGAACGTGCATCAGAGA	67.7°C
P4-1:	CGAGCAGGACCTGGAACCT	67.7°C
P4-2:	TCGACCACTGCAGGTGGA	68.3°C
P4-3:	GCTTCCTCTGCTGCACCA	66.7°C
P4-4:	GGTGTCGAGGTGAGCAGCA	69.1°C
P4-5:	CGACGAGGTCCATCCTGGT	68.6°C
P4-6:	GTGAGGAGCAGGTCCTCCTGT	68.0°C
P4-7:	CTGACACTGGTCGTGGTCGA	68.8°C
P4-8:	CATCTCGACGACCTGCTCCT	67.9°C
P4-9:	ACGAGGTCTCCACTGGTCCA	68.3°C

P4-10: ACTGAGCTGCTCCTCCAGGT 66.5°C

Thermocycling/measurement of fluoresence was performed on an Applied Biosystems ABI Prism 7900HT real-time instrument utilizing the standard cycling parameters:

95°C 10 min;
40 cycles of
95°C 15 sec;
60°C 1 min

All samples were run in duplicate.

Results

Fig. 6 shows the standard curve calculated by the 7900HT system software. The log of the starting copy number was plotted against the measured C_T value. The relationship between C_T and starting copy number was linear in the range from 10 to 10^9 identifier copies.

This standard curve was utilized by the system software to calculate the quantity in the "unknown" samples as shown below.

Table I: Sample A (Shown graphically in Fig. 7)

Sample A: Equimolar ratios	Observed A	Observed B	Expected
FPv2	12539947,00	11977503,00	10000000
P1-1	445841,90	480382,03	1000000
P1-2	884840,70	847478,56	1000000
P1-3	1013073,56	948770,00	1000000
P1-4	764187,94	741304,40	1000000
P1-5	1352874,60	1275155,50	1000000
P1-6	1284075,60	1337928,50	1000000
P1-7	658161,80	747371,56	1000000
P1-8	742187,20	653874,00	1000000
P1-9	824587,75	705785,75	1000000
P1-10	813550,75	836037,90	1000000

P2-1	13145159,00	14482606,00	10000000
P3-1	13263911,00	12773780,00	10000000
P4-1	1430704,80	1472576,80	1000000
P4-2	2681652,00	2481824,80	1000000
P4-3	1933106,80	2085476,40	1000000
P4-4	1359684,40	1364621,40	1000000
P4-5	2206709,80	2065813,60	1000000
P4-6	1652718,10	1873777,20	1000000
P4-7	1468208,10	1416153,00	1000000
P4-8	1664467,50	1581067,00	1000000
P4-9	1462520,60	1594593,80	1000000
P4-10	2020088,20	1912277,40	1000000

Table II: Sample B (Shown graphically in Fig. 8)

Sample B: 2-fold dil.	Observed A	Observed B	Expected
FPv2	4,97E+06	5,05E+06	10000000
P1-1	9955,07	10899,97	9765,625
P1-2	12732,32	13469,12	19531,25
P1-3	25542,8	25419,85	39062,5
P1-4	34748,89	44070,81	78125
P1-5	110881,41	123734,13	156250
P1-6	163687,44	166220,5	312500
P1-7	156993,81	172005,64	625000
P1-8	343176,78	374809,13	1250000
P1-9	646819,44	576151	2500000
P1-10	1,49E+06	1,72E+06	5000000
P2-1	5,19E+06	5,37E+06	10000000
P3-1	5,29E+06	5,09E+06	10000000
P4-1	(no signal)	70223,8	9765,625
P4-2	42103,32	22733,17	19531,25
P4-3	54480,62	39663,62	39062,5
P4-4	51293,07	43950,9	78125
P4-5	137946,95	115027,34	156250

P4-6	174134,64	156442,55	312500
P4-7	316505,78	283856,84	625000
P4-8	737661,44	691296,75	1250000
P4-9	1,42E+06	1,45E+06	2500000
P4-10	3,72E+06	3,52E+06	5000000

The results of the experiments show the possibility of accurately quantification of identifier oligonucleotides down to or even below 10 copies with a 9 fold dynamic range, and reliable relative quantification of the tested codons in various positions in the identifier oligonucleotide.

Example 5 – Codon analysis

Another possibility to analyse codons in identifier oligonucleotides is to use array format with attached probe oligonucleotides.

Six adaptors with the different anti-codon sequences in all three positions were designed. All the adaptors contain a probe binding sequence (20 nucleotides) that allows discrete binding on the microarray. Probe design is known in the art. Adaptors harbouring one to three deletions in the spacing region were used as negative controls to ensure that only the framing region is responsible for the hybridization of the identifier. Thus, the negative controls contain another framing sequence. The identifier oligonucleotide harbours the complementing codon sequence and the position directing framing regions.

Adaptor oligonucleotides

3′CTCATCGGAAGGGCTCGTAACGGTGGGTTTGGGGGCTGGGTTTGGGGCGTGGGTTTGGGCGG-5′

3′TTTGGTAGCTGAGTGCCCTAGGCTGGGTTTGGGCGGTGGGTTTGGGGGCTGGGTTTGGGGCG-5′

3′TAACTGGTTTGACGCCACGCGCGTGGGTTTGGGGCGTGGGTTTGGGCGGTGGGTTTGGGGGC-5′

3′TAATTGAGCTGACGGCGCACGGCTGGGTTTGGGCGTGGGTTTGGGGCTGGGTTTGGGGCG-5′

3′TGTTGCTACTCTGGCCCGAGGCTGGGTTTGGGCTGGGTTTGGGCTGGGTTTGGGGCG-5′

3′ACGGGATAACAACGCAGCCTGGCTGGGTTTGGGTGGGTTTGGGTGGGTTTGGGGCG-5′

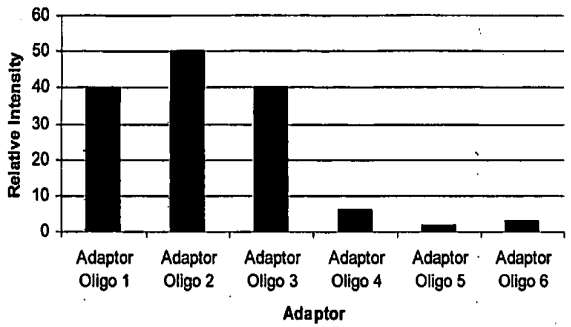
Identifier Oligonucleotide

Biotin-5′GCCACCCAAACCCCCG

GenFlex hybridisation and scanning. Prior to hybridization, the Adaptor mix (100 pM final concentration for each of the adaptor oligonucleotides) in a hybridization buffer (100mM MES, 1 M NaCl, 20 mM EDTA, 0,01% Tween 20, 1x Denhardt's), was heated to 95°C for 5 min and subsequently cooled and maintained at 40°C for 5 min before loading onto the Affymetrix GenFlex probe array cartridge. The probe array was then incubated for 2h at 45°C at constant rotation (60 rpm). The remaining Adaptor mix was removed from the GenFlex cartridge, and replaced with the identifier in a hybridization buffer (100mM MES, 1 M NaCl, 20 mM EDTA, 0,01% Tween 20, 1x Denhardt's). The identifier hybridisation mix was heated to 95°C for 5 min and subsequently cooled and maintained at 40°C for 5 min before loading onto the Affymetrix GenFlex probe array cartridge and hybridised for 2h at 45°C at constant rotation (60 rpm). The washing and staining procedure was performed in the Affymetrix Fluidics Station. The probe array was exposed to 2 washes in 6xSSPE-T at 25°C followed by 12 washes in 0.5xSSPE-T at 40°C. The biotinylated Identifier oligonucleotide was stained with a streptavidin-phycoerythrin conjugate, final concentration 2 µg/µl (Molecular Probes, Eugene, OR) in 6xSSPE-T for 10 min at 25°C followed by 6 washes in 6xSSPE-T at 25°C.

The probe arrays were scanned at 560 nm using a confocal laser-scanning microscope with an argon ion laser as the excitation source (Hewlett Packard GeneArray Scanner G2500A). The readings from the quantitative scanning were analysed by the Affymetrix Gene Expression Analysis Software. The results are depicted in Scheme 1.

Scheme 1:



The Array analysis shows that the codons including the framing regions are able to distinguish between the different probe oligonucleotides. The designed probes will only detect codons with the correct framing region allowing distinguishing first of the right codon and secondly as to which position the codon is positioned. Only one deletion in both framing regions reduces significantly the hybridization of the identifier. Thus, the framing sequence may be used to obtain information about the position of a specific codon and the point in the reaction history when a given reaction of a chemical entity has occurred.

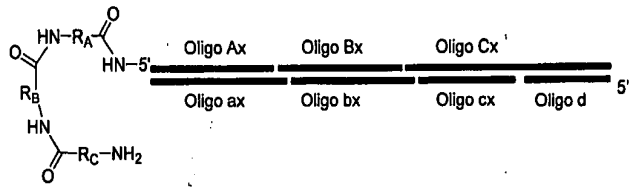
The information obtained in this example using either QPCR or array codon analysis as example can be used to generate a new more focused library. The signal from the QPCR analysis or the array analysis can directly be used to combine preferable chemical entities.

EXAMPLE 6. Generation of a second-generation library.

The information obtained from a codon analysis performed according to the principles described in Examples 4 or 5 can be utilized for assembly a new more focused library. Sequence information can also be used to design a second-generation library with reduced diversity. This example illustrates how sequence data can be utilized to make a more focused library with the enriched chemical entities. Identical strategy can be based on the codon analysis methods described in Examples 4 or 5.

A 700-member library was generated composing of 4 x 25 x 7 chemical entities. The library generation protocol is described below with the sequence information and chemical entity structure.

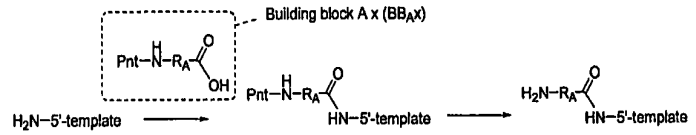
General arrangement of each complex composed of display molecule and identifier oligonucleotide in the library generation:



Specific codons in each oligo (Ax, Bx, Cx) was used and can be designed by using a specific nucleotide sequence for each chemical entity. In this particular setup, two complementary oligonucleotides (e.g. oligo Ax and oligo ax) containing a particular codon are allow to hybridize before the ligation step. The ligation of each codon oligonucleotide in each position is ligated with that attachment of the encoded chemical entity.

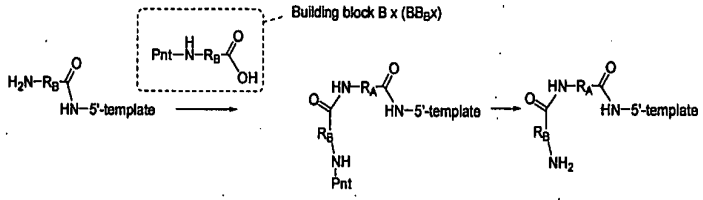
Overview of the library generation procedure:

First round of library generation (round A) :

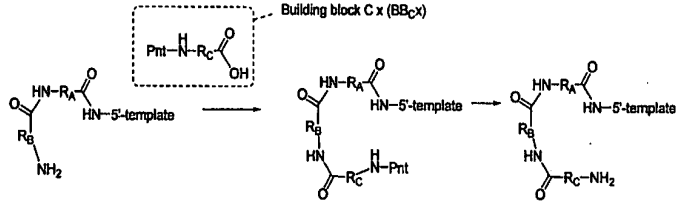


"Pnt" corresponds to pentenoyl – an amine protecting group. "R" can by any molecule fragment. The chemical used in library generation comprise a primary (shown) or a secondary amine.

Second round of library generation (round B) :



Third round of library generation (round C):



General procedure : Library generation, selection and mismatch subsequent selection

First round of library generation (round A) :

First oligonucleotides of the A series are each modified by adding to each type of oligo a small molecule building block (BB_Ax) to the 5' amine forming an amide bond. After this step the identifier is comprised of oligo Ax.

Second round of library generation (round B) :

4 nmol of a mixture of different modified A oligos are then split into a number tubes corresponding to the number of different building blocks to be used in round B. 190 pmol Oligo a and 2 µl heering DNA is added to each tube and the DNA material in each tube is lyophilized. The lyophilized DNA is then redissolved in 50 µl water and purified by spining through Biospin P-6 columns (Biorad) equilibrated with water.

Addition of building block

The DNA material in each tube is again lyophilized and redissolved in 2 µl 100 mM Na-borate pH 8.0/100 mM sulfo N-hydroxy succinimide (sNHS). For each tube 10 µl building block BB_Bx (100 mM in dimethyl sulfoxide [DMSO]) is preactivated by mixing with 10 µl 1-Ethyl-3-(3-dimethylaminopropyl)-carbodiimide (EDC) (90 mM in dimethylformamide [DMF]) and incubating at 30°C for 30 min. 3 µl of this preactivated mixture is then mixed with the 2 µl in each tube and allowed to react 45 min at 30 °C. Then an additional 3 µl freshly preactivated BB is added and the reaction is allowed to proceed for 45 min at 30 °C. The resulting mixture is then purified by spinning through Bio-Rad P6 DG (Desalting gel).

Addition of codon oligonucleotide

The DNA material is then lyophilized and redissolved in 10 µl water containing 200 pmol oligo Bx (eg. B1) and the corresponding oligo bx (eg. b1). This is done so that the codon in oligo Bx identifies the BB_Bx added to the DNA identifier. 10 units of T4 DNA ligase (Promega) and 1.2 µl T4 DNA ligase buffer is then added to each tube and the mixture is incubated at 20°C for 1 hour. The DNAn identifier linked to the small molecules now comprises an Ax oligo with a Bx oligo ligated to its 3' end. The reactions are then pooled, an appropiate volume of water is allowed to evaporate and the remaining sample is purified by spining through Biospin P-6 columns (Biorad) equilibrated with water.

Removal of building block protecting group

The pooled sample (~ 50 µl) is adjusted to 10 mM Na-acetate (pH 5). 0.25 volumes of 25 mM Iodine in tetrahydrofuran/water (1:1) is added and the sample is incubate at 37 °C for 2h. The reaction is then quenched by addition of 2 µl of 1M $Na_2S_2O_3$ and incubation at room temperature for 5 min. The complexes are then purified by spining through Biospin P-6 columns (Biorad) equilibrated with water

To remove sulphonamide protecting groups, the sample is adjusted to 50 µl 100 mM sodium borate pH 8.5 and 20 µl 500 mM 4-methoxy thiophenol (in acetonitrile) is added and the reaction is incubated at 25°C overnight. Then the complexes are purified by spinning through Biospin P-6 columns (Biorad) equilibrated with water and then lyophilized.

Third round of library generation (round C) :

The samples are dissolved in 175 µl 100 mM Na-borate pH 8.0 and distributed into 25 wells (7 µl / well). 2 µl 100 mM BB_Cx in water/DMSO and 1 µl of 250 mM DMT-MM is added to each reaction and incubated at 30 °C overnigth. Water is added to 50 µl and the reactions are then spin purified using Bio-Rad P6 DG (Desalting gel) and subsequently water is allowed to evaporate so that the final volume is 10 µl.

Addition of building block

The DNA material is then lyophilized and redissolved in 10 µl water containing 200 pmol oligo Cx (eg. C1) and the corresponding oligo cx (eg. c1). This is done so that the codon in oligo Cx corresponds to the BB_Cx added to the DNA identifier. 10 units of T4 DNA ligase (Promega) and 1.2 µl T4 DNA ligase buffer is then added to each tube and incubated at 20ºC for 1 hour. The DNAn identifier linked to the small molecules now comprises and Ax oligo with a Bx ligated to its 3' end and a Cx oligo ligated to the 3' end of the Bx oligo. The reactions are then pooled, the pooled sample volume is reduced by evaporation and the sample is purified by spining through Biospin P-6 columns (Biorad) equilibrated with water. The pooled sample (~ 50 µl) is adjusted to 10 mM Na-acetate (pH 5). 0.25 volumes of 25 mM Iodine in tetrahydrofuran/water (1:1) is added and the sample is incubate at 37 °C for 2h. The reaction is then quenched by addition of 2 µl of 1M $Na_2S_2O_3$ and incubation at RT for 5 min. Then the DNA identifiers (carrying small molecules) are purified by spinning through Biospin P-6 columns (Biorad) equilibrated with water and then lyophilized.

Final deprotection step

Some building blocks contain methyl esters that are deprotected to acids by dissolving the pooled sample in 5 µl 20 mM NaOH, heating to 80 °C for 10 minutes and adding 5 µl of 20 mM HCl.

Final extension step

To ensure that the DNA identifiers are double stranded prior to selection oligo d is extended along the identifier by adding to the sample 10 µl of 5 X sequenase EX-buffer [100 mM Hepes, pH 7.5, 50 mM $MgCl_2$, 750 mM NaCl] and 4000 pmol oligo d. Annealing is performed by heating to 80ºC and cooling to 20 °C. To the sample is then added 500 µL dNTP , water to 50 µl and 39 units of Sequenase version 2.0 (USB). The reaction is incubated at 37ºC for 1 hour.

Selection

This library is subjected to selection, whereby binders to the selection target are enriched.

Maxisorp ELISA wells (NUNC A/S, Denmark) were coated with each 100 µL 2µg/mL integrin αVβ3 in PBS buffer [2.8 mM NaH_2PO_4, 7.2 mM Na_2HPO_4, 0.15 M NaCl, pH 7.2] overnight at 4°C. Then the integrin solution was substituted for 200 µl blocking buffer [TBS, 0.05% Tween 20 (Sigma P-9416), 1% bovine serum alnumin (Sigma A-7030), 1 mM $MnCl_2$] which was left on for 3 hours at room temperature. Then the wells were washed 10 times with blocking buffer and the encoded library was added to the wells after diluting it 100 times with blocking buffer. Following 2 hours incubation at room temperature the wells were washed 10 times with blocking buffer. After the final wash the wells were cleared of wash buffer and subsequently inverted and exposed to UV light at 300-350 nm for 30 seconds using a trans-illuminator set at 70% power. Then 100 µl blocking buffer without Tween-20 was immediately added to each well, the wells were shaken for 30 seconds, and the solutions containing eluted identifiers were removed for PCR amplification.

Cloning

A TOPO-TA (Invitrogen) ligation reaction is assembled with 4 µl PCR product, 1 µl salt solution (Invitrogen) and 1 µl vector. Water is added to 6 µl. The reaction is then incubated at RT for 30 min. Heat-shock competent TOP10 E.coli cells are then thawed on ice and 5 µl of the ligation reaction is added to the thawed cells. The cells are then incubated 30 min on ice, heatshocked in 42ºC water for 30 sec, and then put on ice again. 250 µl of growth medium is added to the cells and they are incubated 1 h at 37ºC. The medium containting cells is then spread on a growth plate containing 100 µg / ml ampicillin and incubated at 37ºC for 16 hours.

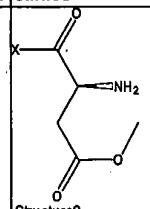
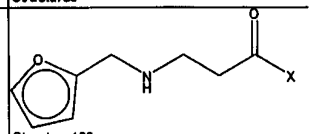
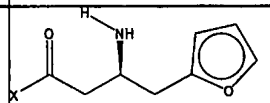
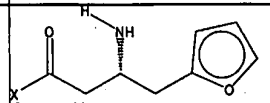
Sequencing

Individual *E.coli* clones are then picked and transferred to PCR wells containing 50 µl water. These 50 µl were incubated at 94ºC for 5 minutes and used in a 20 µl in a 25 µl PCR reaction with 5 pmol of each TOPO primer M13 forward & M13 reverse and Ready-To-Go PCR beads (Amersham Biosciences). The following PCR profile is used: 94ºC 2 min, then 30 x (94ºC 4 sec, 50ºC 30 sec, 72ºC 1 min) then 72ºC 10 min. Primers and nucleotides are then degraded by adding 1 µl 1:1 EXO/SAP mixture (USB corp.) to 2 µl PCR product and incubating at 37ºC for 15 min and then 80ºC for 15 min

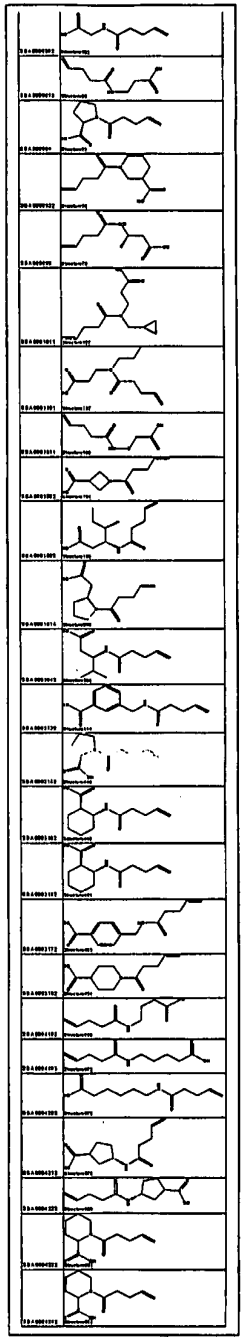
to heat-inactivate the enzymes. 5 pmol T7 primer is added and water is added to 12 µl. Then 8 µl DYEnamic ET cycle sequencing Terminator Mix (Applied biosystems) is added to each well. A thermocycling profile of 30 x (95°C 20 sec, 50°C 15 sec, 60°C 1 min) is then run. Then 10 µl water is added to each well and sequencing reactions are purified using seq96 spinplates (Amersham Biosciences). Reactions are then run on a MegaBace capillary electrophoresis instrument (Molecular Dynamics) using injection parameters 2 kV, 50 sec and run parameters: 9 kV 45 min and analyzed using Contig Express software (Informax).

The chemical entities used in each position are shown below.

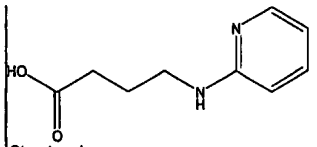
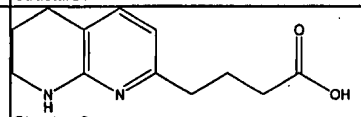
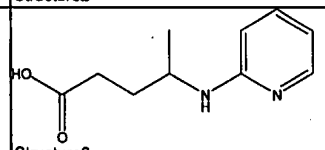
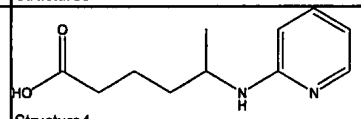
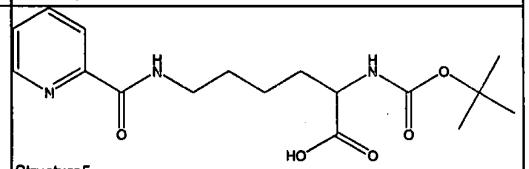
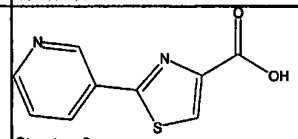
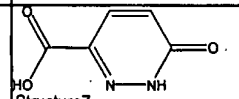
Position 1

Building Block	Smiles
BB-A-000098	 Structure9
BB-A-000112	 Structure123
BB-A-000282	 Structure81
BB-A-000283	 Structure82

Position 2



Position 3

BBA0000531	Structure1 
BBA0001006	Structure2 
BBA0001391	Structure3 
BBA0001401	Structure4 
BBA0008312	Structure5 
BBA0008512	Structure6 
BBA0008812	Structure7 

After the selection as described above, the codons in the identifier oligonucleotides were analysed. Before the analysis, the identifier oligonucleotides were amplified using the constant flanking regions and the amplified material was used in the identifier sequence analysis.

A sequence codon analysis of the selected codons showed a bias for specific chemical entities. They are listed in the table below. For instance, in position 1 chemical entity 98 was seem 47 times (out of 51 sequences, 92%, compare to 25 % before the selection) and chemical entity 99 was seen 14 times (out 51 sequences, 27%, compare to 4% before selection) and chemical entity 53 was seen 35 times (out of 51 sequences, 68%, compare to 14% before selection).

The chemical entities listed in the table below can then be used to generate a new and more focused library.

Oligo(-s)	Count	pos 1	pos 2	pos 3
BB-A-000098	47	98		
BB-A-000282	4	282		
[redacted]				
[redacted]				
BBA0004242	6		424	
BBA0004182	5		418	
BBA0001101	2			
BBA0003172	2			
BBA0004212	2			
BBA0004232	2			
BBA000064	1			
BBA0001011	1			
BBA0003132	1			
BBA0003142	1			
BBA0003152	1			
[redacted]				
[redacted]				
BBA0001006	4			100
BBA0008512	2			
BBA0008312	1			

The new focused library with the selected chemical entities can be selected against the target and the outcome from the selection can be analysed. The most abundant binders will be the combination between the chemical entities 98-99-53 and the second most abundant binder is 98-158-53 as shown below.

Oligo(-s)	Count	pos 1	pos 2	pos 3
BB-A-000098 BBA000099 BBA0000531	11	98		53
BB-A-000098 BBA0001582 BBA0000531	7	98	158	63
BB-A-000098 BBA0004242 BBA0000531	4	98	424	53
BB-A-000098 BBA0001582 BBA0001391	3	98	158	139
BB-A-000098 BBA0004182 BBA0000531	3	98	418	53
BB-A-000098 BBA000099 BBA0001391	2	98		139
BB-A-000098 BBA0001582 BBA0001006	2	98	158	100

This example exemplifies the possibility to reduce the library diversity by using the enriched chemical entities in a new library and perform another round of selection on the chosen chemical entities.

Example 7

The following experiment illustrates the principle of chemical entity (also termed building block herein) evolution through multiple rounds of library generation and selection. The experiment is not intended to limit the scope of the current invention.

Libraries were assembled by the combination of building blocks (BB) each of which was encoded by an oligonucleotide (oligo). Some of the building blocks carried an amine functional group and a carboxylic acid functional group. The building block amine was protected by *N*-pentenoylation and deprotected by iodine treatment prior to the reaction of the following building block. Oligonucleotide 1 (Oligo1) carried an amine functional group to allow reaction with the building block1's carboxylic acid and oligonucleotides are optionally derivatized by phosphorylation to allow ligation. Oligonucleotide3 (oligo3) also comprised a primer region for PCR amplification. EDC/NHS, EDC/sulfoNHS or DMTMM was used as coupling reagents.

The following scheme describes the split and mix assembly of the libraries:

i.) n times [BB1 + Oligo1 → BB1-Oligo1] in separate wells

* Optionally purify product

ii.) mix all n wells into one tube

iii.) split product of ii.) into m separate wells

iv.) m times [BB2 + BB1-Oligo1 + Oligo2 → BB2-BB1-Oligo1-Oligo2] in separate wells

* Optionally purify product

v.) mix all m wells into one tube

vi.) split product of v.) into p separate wells

vii.) p times [BB3 + BB2-BB1-Oligo1-Oligo2 + Oligo3 → BB3-BB2-BB1-Oligo1-Oligo2-Oligo3] in separate wells

* Optionally purify product

viii.) mix all p wells into one tube

ix.) Selection was performed and binders isolated

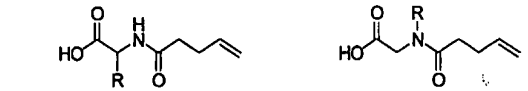
x.) PCR of DNA and sequencing

xi.) Analyse for building block abundancy and full sequence information

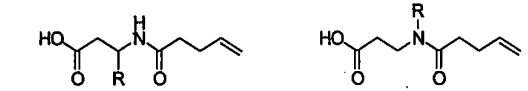
Building block abundances analysis may be done by QPCR or by sequencing full sequences and then analyzing for the abundance of individual building blocks.

The following types of building blocks were used, wherein R describes a group which is varied for different building blocks:

Building block types used in position 1, 2 and 3



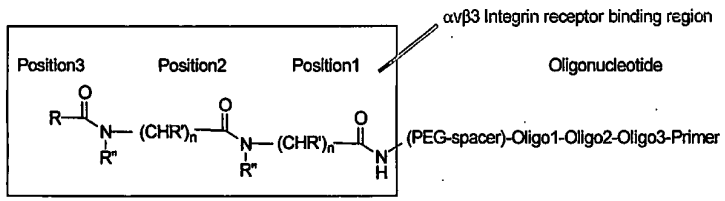
Both (*R*) and (*S*) isomers were for some R-groups used



Both (*R*) and (*S*) isomers were for some R-groups used

Building block types which were only used in position 3

The overall process leads to molecules of the following structure, where the oligonucleotide was double stranded.



R' = H or R (as indicated for building blocks)
R'' = H or R (as indicated for building blocks)
R = H or R (as indicated for building blocks)
n = 1-2

The oligonucleotide was made double stranded by the use of double stranded Oligo's 1, 2 and 3 with an overhang to allow ligation of both strands.

Summary of the experimental outcome:

Two libraries of 61,875 members (Library 1 and 2) were generated as described in example 6 above and selected for binders of the Integrin αvβ3 receptor separately.
The libraries were generated with 99 different building blocks in position 1, 25 different building blocks in position 2 and 25 different building blocks in position 3.
The identified sequences were then analyzed for the abundances of building blocks at each position in the sequence. The most abundant building blocks at each position from the two libraries 1 and 2 were then used again to generate a new and smaller library of 1,365 members, which was selected for binders of the Integrin αvβ3 receptor.
The library was generated with 7 different building blocks in position 1, 13 different building blocks in position 2 and 15 different building blocks in position 3.
In the tables below, each of the building block numbers identify one specific building block or in two instances (library 1) a mixture of three different building blocks. The same numbers are used for each building block in all libraries, however the oligonucleotide used to identify each building block may not necessarily be the same between libraries to avoid potential problems of cross contamination.

The following tables describes the codon sequences and corresponding building blocks used. The codon is only indicated for one of the strands.

Library 1, Position 1

Codon no.	Codon sequence ID	Building Block ID
1	TGTTC	BBA000092
2	CGAGC	BBA000354
3	GGATA	BBA000085
4	CGCTG	BBA000086
5	GTTAT	BBA000098
6	AGTGC	BBA000099
7	ACCTG	BBA000089
8	CTGGT	BBA000090
9	TAGGA	BBA000087
10	ACTCA	BBA000088
11	CTTAC	BBA000153
12	CGCAC	BBA000154
13	TCGCG	BBA000059

14	CGGAT	BBA000152
15	GAGAT	BBA000101
16	TGTAG	BBA000110
17	GTGTT	BBA000112
18	AGATG	BBA000113
19	ATCCT	BBA000114
20	TTGCT	BBA000286
21	ACGTA	BBA000123
22	ATCAC	BBA000124
23	TATCC	BBA000155
24	GGAAG	BBA000156
25	CGGTC	BBA000158
26	TGCTT	BBA000159
27	TTAGC	BBA000160
28	GCTGA	BBA000161
29	GAACG	BBA000162
30	CATGG	BBA000163
31	TGGTA	BBA000165
32	TCAAG	BBA000166
33	ATCGA	BBA000167
34	ATGCA	BBA000168
35	ACTAG	BBA000169
36	TACCT	BBA000170
37	TACGA	BBA000171
38	CTTCA	BBA000172
39	CTCTT	BBA000173
40	TCATC	BBA000174
41	ATTCC	BBA000175
42	CGACG	BBA000176
43	CCTGT	BBA000177
44	CCTTC	BBA000178
45	ACACC	BBA000179
46	TAACA	BBA000180
47	TAACA	BBA000098
48	CCAGG	BBA000181
49	ATGTC	BBA000182
50	GAGGA	BBA000183

51	GGTCA	BBA000184
52	GACTT	BBA000185
53	GGTGG	BBA000186
54	CAACT	BBA000190
55	ATGAG	BBA000195
56	TCTGC	BBA000196
57	ATAGG	BBA000197
58	CTACC	BBA000198
59	AAGTG	BBA000201
60	TCCAA	BBA000202
61	GCTCT	BBA000203
62	GGAGT	BBA000204
63	AATCG	BBA000205
64	AAGCT	BBA000206
65	CCGAA	BBA000207
66	TTTGT	BBA000208
67	CCGTG	BBA000209
68	TTTCG	BBA000210
69	TGAGG	BBA000211
70	GTTGC	BBA000212
71	AACTA	BBA000112
72	AACTA	BBA000280
73	CCTCG	BBA000281
74	AGCAA	BBA000282
75	TTCCA	BBA000313
76	AGACT	BBA000314
77	AGGTT	BBA000315
78	GCGTC	BBA000316
79	AACGT	BBA000317
80	CAAGA	BBA000287
81	AGAGA	BBA000419
82	GTACT	BBA000420
83	TAGAG	BBA000421
84	ACGAT	BBA000422
85	GACCA	BBA000200
86	TCGTT	BBA000194
87	GTCTC	BBA000427

88	CAGCA	BBA000428
89	TAGTC	BBA000199
90	GGGTG	BBA000187
91	CTCAG	BBA000191
92	AGAAC	BBA000284
93	GCGAG	BBA000458
94	GATGT	BBA000459
95	TCACT	BBA000461
96	CGTCT	OBA000610
97	AGCTC	OBA000611
98	CACTC	OBA000609
99	CAGTT	OBA000615

Library 1, Position 2

Codon no.	Codon sequence ID	Building Block ID
1	AGTACGAACGTGCATCAGAG	BBA000098
2	TAGTCTCCTCCACTTCCATG	BBA000099
3	TACATCGTTCCAGACTACCG	BBA000085
4	TCCAGTGCAAGACTGAACAG	BBA000153
5	AGCATCACTACTCTGTCTGG	BBA000206
6	TCTTGTCAACCTTCCATGCG	BBA000200
7	AAGGACGTTCCTAGTAGGTG	BBA000208
8	GGAACCATCAAGATCCTGAG	BBA000091
9	ATCTCTGACGAGATCCAAGG	BBA000090
10	TCAAGGTTGGTGGTGTACTG	BBA000092
11	TCGAACTTGTTGCTTCCTCG	BBA000123
12	CTGAGTGTGTAGTACCAACG	BBA000156
13	ATCTTGGTTGTTCTCCTGCG	BBA000163
14	TAGTAGCTTGGAGTAGACCG	BBA000197
15	TTCACTCCATGCAGCATGTG	BBA000083
16	ACGATGGTGATCGATCAACG	BBA000181
17	TTCAGTGCTTGAGCTACCTG	BBA000152
18	TTGGACTCTTCTTGCACCAG	BBA000088
19	TCAACCAACTGGTTCTTGGG	BBA000100

20	TAGTACTCTACACTGCTGCG	BBA00087/101/196
21	TACACCATGACTTGCAGACG	BBA00087/101/196
22	GCATCTTGAGTCGTTGAACG	BBA000059
23	GACTCATCTCACTGGAGTTG	BBA000124
24	TCCAGCTTCTAGGAAGACAG	BBA000160
25	CTTCTTGAGTGCACTAGCAG	BBA000201

Library 1, Position 3

Codon no.	Codon sequence ID	Building Block ID
1	CGAGCAGGACCTGGAACCTGGTGC	BBA000098
2	CTCGACCACTGCAGGTGGAGCTCC	BBA000099
3	CGTGCTTCCTCTGCTGCACCACCG	BBA000085
4	CCTGGTGTCGAGGTGAGCAGCAGC	BBA000153
5	CTCGACGAGGTCCATCCTGGTCGC	BBA000206
6	CGTGAGGAGCAGGTCCTCCTGTCG	BBA000200
7	CCTGACACTGGTCGTGGTCGAGGC	BBA000208
8	CCATCTCGACGACCTGCTCCTGGG	BBA000091
9	CCACGAGGTCTCCACTGGTCCAGG	BBA000090
10	CCACTGAGCTGCTCCTCCAGGTGG	BBA000092
11	CCTCCTGTCCTGCACGTCCATCCG	BBA000123
12	CAGCACCTGGAGGTAGGACCACGG	BBA000156
13	CGACCAGACGAGGACCAGGTAGGC	BBA000163
14	CCAGGTTCGAGGACCTCGTCAGCC	BBA000197
15	CGAGCACGAGGAGCACGTGTCCAG	BBA000100
16	CCACGTCCACAGGTGCACCAGGTG	BBA000181
17	CCTGGTGCTCCACGACGTGCTTCG	BBA000152
18	CACGTGACGACCTGGTCAGGTGGG	BBA000088
19	CGTAGCTCGTGCTGGTCCTCCTGG	BBA000101
20	CGACGACCACCACCTTGGACACCC	BBA000196
21	CCTACGTCGTGCTCACGTCCTGCC	BBA00087
22	CGACGACAGCTAGGAGGAGGTGGG	BBA000083
23	CTGGTGGAGCTGCACGAGCACAGC	BBA000059
24	CAGGACTGGACGACGACCAGGTCG	BBA000124
25	CGATGCTGCAGACGACCAGCACCC	BBA000160

Library 2, Position 1

Codon no.	Codon sequence ID	Building Block ID
1	TGTTC	BBA000092
2	CGAGC	BBA000354
3	GGATA	BBA000085
4	CGCTG	BBA000086
5	GTTAT	BBA000098
6	AGTGC	BBA000099
7	ACCTG	BBA000089
8	CTGGT	BBA000090
9	TAGGA	BBA000087
10	ACTCA	BBA000088
11	CTTAC	BBA000153
12	CGCAC	BBA000154
13	TCGCG	BBA000059
14	CGGAT	BBA000152
15	GAGAT	BBA000101
16	TGTAG	BBA000110
17	GTGTT	BBA000112
18	AGATG	BBA000113
19	ATCCT	BBA000114
20	TTGCT	BBA000286
21	ACGTA	BBA000123
22	ATCAC	BBA000124
23	TATCC	BBA000155
24	GGAAG	BBA000156
25	CGGTC	BBA000158
26	TGCTT	BBA000159
27	TTAGC	BBA000160
28	GCTGA	BBA000161
29	GAACG	BBA000162
30	CATGG	BBA000163
31	TGGTA	BBA000165

32	TCAAG	BBA000166
33	ATCGA	BBA000167
34	ATGCA	BBA000168
35	ACTAG	BBA000169
36	TACCT	BBA000170
37	TACGA	BBA000171
38	CTTCA	BBA000172
39	CTCTT	BBA000173
40	TCATC	BBA000174
41	ATTCC	BBA000175
42	CGACG	BBA000176
43	CCTGT	BBA000177
44	CCTTC	BBA000178
45	ACACC	BBA000179
46	TAACA	BBA000180
47	TAACA	BBA000098
48	CCAGG	BBA000181
49	ATGTC	BBA000182
50	GAGGA	BBA000183
51	GGTCA	BBA000184
52	GACTT	BBA000185
53	GGTGG	BBA000186
54	CAACT	BBA000190
55	ATGAG	BBA000195
56	TCTGC	BBA000196
57	ATAGG	BBA000197
58	CTACC	BBA000198
59	AAGTG	BBA000201
60	TCCAA	BBA000202
61	GCTCT	BBA000203
62	GGAGT	BBA000204
63	AATCG	BBA000205
64	AAGCT	BBA000206
65	CCGAA	BBA000207
66	TTTGT	BBA000208
67	CCGTG	BBA000209
68	TTTCG	BBA000210

69	TGAGG	BBA000211
70	GTTGC	BBA000212
71	AACTA	BBA000112
72	AACTA	BBA000280
73	CCTCG	BBA000281
74	AGCAA	BBA000282
75	TTCCA	BBA000313
76	AGACT	BBA000314
77	AGGTT	BBA000315
78	GCGTC	BBA000316
79	AACGT	BBA000317
80	CAAGA	BBA000287
81	AGAGA	BBA000419
82	GTACT	BBA000420
83	TAGAG	BBA000421
84	ACGAT	BBA000422
85	GACCA	BBA000200
86	TCGTT	BBA000194
87	GTCTC	BBA000427
88	CAGCA	BBA000428
89	TAGTC	BBA000199
90	GGGTG	BBA000187
91	CTCAG	BBA000191
92	AGAAC	BBA000284
93	GCGAG	BBA000458
94	GATGT	BBA000459
95	TCACT	BBA000461
96	CGTCT	OBA000610
97	AGCTC	OBA000611
98	CACTC	OBA000609
99	CAGTT	OBA000615

Library 2, Position 2

Codon no.	Codon sequence ID	Building Block ID

1	AGTACGAACGTGCATCAGAG	BBA000059
2	TAGTCTCCTCCACTTCCATG	BBA000085
3	TACATCGTTCCAGACTACCG	BBA000098
4	TCCAGTGCAAGACTGAACAG	BBA000099
5	AGCATCACTACTCTGTCTGG	BBA000101
6	TCTTGTCAACCTTCCATGCG	BBA000110
7	AAGGACGTTCCTAGTAGGTG	BBA000113
8	GGAACCATCAAGATCCTGAG	BBA000114
9	ATCTCTGACGAGATCCAAGG	BBA000123
10	TCAAGGTTGGTGGTGTACTG	BBA000124
11	TCGAACTTGTTGCTTCCTCG	BBA000152
12	CTGAGTGTGTAGTACCAACG	BBA000158
13	ATCTTGGTTGTTCTCCTGCG	BBA000160
14	TAGTAGCTTGGAGTAGACCG	BBA000161
15	TTCACTCCATGCAGCATGTG	BBA000167
16	ACGATGGTGATCGATCAACG	BBA000176
17	TTCAGTGCTTGAGCTACCTG	BBA000181
18	TTGGACTCTTCTTGCACCAG	BBA000313
19	TCAACCAACTGGTTCTTGGG	BBA000314
20	TAGTACTCTACACTGCTGCG	BBA000315
21	TACACCATGACTTGCAGACG	BBA000316
22	GCATCTTGAGTCGTTGAACG	BBA000317
23	GACTCATCTCACTGGAGTTG	BBA000420
24	TCCAGCTTCTAGGAAGACAG	BBA000421
25	CTTCTTGAGTGCACTAGCAG	BBA000422

Library 2, Position 3

Codon no.	Codon sequence ID	Building Block ID
1	CGAGCAGGACCTGGAACCTGGTGC	BBA000052
2	CTCGACCACTGCAGGTGGAGCTCC	BBA000053
3	CGTGCTTCCTCTGCTGCACCACCG	BBA000054
4	CCTGGTGTCGAGGTGAGCAGCAGC	BBA000056
5	CTCGACGAGGTCCATCCTGGTCGC	BBA000057
6	CGTGAGGAGCAGGTCCTCCTGTCG	BBA000058

7	CCTGACACTGGTCGTGGTCGAGGC	BBA000062
8	CCATCTCGACGACCTGCTCCTGGG	BBA000139
9	CCACGAGGTCTCCACTGGTCCAGG	BBA000140
10	CCACTGAGCTGCTCCTCCAGGTGG	BBA000100
11	CCTCCTGTCCTGCACGTCCATCCG	BBA000059
12	CAGCACCTGGAGGTAGGACCACGG	BBA000085
13	CGACCAGACGAGGACCAGGTAGGC	BBA000098
14	CCAGGTTCGAGGACCTCGTCAGCC	BBA000099
15	CGAGCACGAGGAGCACGTGTCCAG	BBA000101
16	CCACGTCCACAGGTGCACCAGGTG	BBA000110
17	CCTGGTGCTCCACGACGTGCTTCG	BBA000113
18	CACGTGACGACCTGGTCAGGTGGG	BBA000114
19	CGTAGCTCGTGCTGGTCCTCCTGG	BBA000123
20	CGACGACCACCACCTTGGACACCC	BBA000124
21	CCTACGTCGTGCTCACGTCCTGCC	BBA000152
22	CGACGACAGCTAGGAGGAGGTGGG	BBA000158
23	CTGGTGGAGCTGCACGAGCACAGC	BBA000160
24	CAGGACTGGACGACGACCAGGTCG	BBA000161
25	CGATGCTGCAGACGACCAGCACCC	BBA000167

Library 3, Position 1

Codon no.	Codon sequence ID	Building Block ID	More abundant in position 1 in library no.
1	TGTTC	BBA000092	1
2	ACTCA	BBA000088	1
3	CTTAC	BBA000153	1 and 2
4	CGGAT	BBA000152	1
5	ATTCC	BBA000175	1 and 2
6	GTCTC	BBA000427	1
7	ACAGT	BBA000098	1 and 2

Library 3, Position 2

Codon no.	Codon sequence ID	Building Block ID	More abundant in in positio

			2 in library no.
1	6CACAAGTACGAACGTGCATCAGAG	BBA000059	1
2	6CACATAGTCTCCTCCACTTCCATG	BBA000083	1
3	6CACATACATCGTTCCAGACTACCG	BBA000085	2
4	6CACATCCAGTGCAAGACTGAACAG	BBA000088	1
5	6CACAAGCATCACTACTCTGTCTGG	BBA000090	1
6	6CACATCTTGTCAACCTTCCATGCG	BBA000099	1 and 2
7	6CACAAAGGACGTTCCTAGTAGGTG	BBA000110	
8	6CACAGGAACCATCAAGATCCTGAG	BBA000114	2
9	6CACAATCTCTGACGAGATCCAAGG	BBA000152	2
10	6CACATCAAGGTTGGTGGTGTACTG	BBA000160	2
11	6CACATCGAACTTGTTGCTTCCTCG	BBA000200	1
12	6CACACTGAGTGTGTAGTACCAACG	BBA000201	1
13	6CACAATCTTGGTTGTTCTCCTGCG	BBA000422	2

Library 3, Position 3

Codon no.	Codon sequence ID	Building Block ID	More abunda in in po tion 3 i library
1	6AGGACGAGCAGGACCTGGAACCTGGTGCGTTCCTCCACCACGTCTCCG	BBA000053	2
2	6AGGACTCGACCACTGCAGGTGGAGCTCCGTTCCTCCACCACGTCTCCG	BBA000085	1
3	6AGGACGTGCTTCCTCTGCTGCACCACCGGTTCCTCCACCACGTCTCCG	BBA000087	1
4	6AGGACCTGGTGTCGAGGTGAGCAGCAGCGTTCCTCCACCACGTCTCCG	BBA000090	1
5	6AGGACTCGACGAGGTCCATCCTGGTCGCGTTCCTCCACCACGTCTCCG	BBA000091	1
6	6AGGACGTGAGGAGCAGGTCCTCCTGTCGGTTCCTCCACCACGTCTCCG	BBA000098	1
7	6AGGACCTGACACTGGTCGTGGTCGAGGCGTTCCTCCACCACGTCTCCG	BBA000100	1 and 2
8	6AGGACCATCTCGACGACCTGCTCCTGGGGTTCCTCCACCACGTCTCCG	BBA000139	2
9	6AGGACCACGAGGTCTCCACTGGTCCAGGGTTCCTCCACCACGTCTCCG	BBA000140	2
10	6AGGACCACTGAGCTGCTCCTCCAGGTGGGTTCCTCCACCACGTCTCCG	BBA000152	
11	6AGGACCTCCTGTCCTGCACGTCCATCCGGTTCCTCCACCACGTCTCCG	BBA000153	1
12	6AGGACAGCACCTGGAGGTAGGACCACGGGTTCCTCCACCACGTCTCCG	BBA000161	
13	6AGGACGACCAGACGAGGACCAGGTAGGCGTTCCTCCACCACGTCTCCG	BBA000167	2
14	6AGGACCAGGTTCGAGGACCTCGTCAGCCGTTCCTCCACCACGTCTCCG	BBA000197	1

15	6AGGACGAGCACGAGGAGCACGTGTCCAGGTTCCTCCACCACGTCTCCG	BBA000200	1

A subset of the isolated sequences from the library post selection was analysed:

(1)
GGCAGCACAGTCGTCGCACATACATCGTTCCAGACTACCGAGGACCTGACACTGGTCGTGGTCGAGGCGTTCCT
(2) GGCAGCACAGT
CGTCGCTACATGCTTGTCAACCTTCCATGCGAGTACCTTACACTGGTTCGTGGTCGAGGCGTTCCT
(3)
GGCAGCCGGAT423CGTCGCACATCTTGTCAACCTTCCATGCGAGGACCTGACACTGGTCGTGGTCGAGGCGTTC
CT
(4)
GGCAGCCTTACGTCGCACAATTCTCTGACAGAAATCCAACGGAGGACCTGACACGTGCGTCGTGGCTCGATGCGT
TCCTC
(5)
GGCAGCACAGTCGTCGCACATCATTGTACAAACCTTCCATGCGAGGACCATCTCGACGACCTGCTCCTGGGGTNC
CTC
(6)
GGCAGCACAGTCGTCGCACATCTTGTCAACCTTCCATGCGAGGACCATCTCGACGACCTGCTCCTGGGGTTCCTC
(7)
GGCAGCACAGTCGTCGCACATCTTGTCAACCTTCCATGCGAGGACCATCTCGACGACCTGCTCCTGGGGTTCCTC
(8)
GGCAGCACAGTCGTCGCACATCTTGTCAACCTTCCATGCGAGGACCATCTCGACGACCTGCTCCTGGGGTTCCTC
(9)
GGCAGCACAGTCGTCGCACATCTTGTCAACCTTCCATGCGAGGACCATCTCGACGACCTGCTCCTGGGGTTCCTC
(10)
GGCAGCACAGTCGTCGCACATCTTGTCAACCTTCCATGCGAGGACCATCTCGACGAGCTGCTCCTGGGGTTCCTC
(11)
GGCAGCACTAGATCGTCGCACATCTTGTCAACCTTCCATGCGAGGACCATCTTCGACTGANCTGCCTCCTGTGGGC
TTCCTC
(12) GGCAGCACAGAT
CGTCGCACATCTTGTCAACCTTCCATGCGAGGACCATCTCGACGANCTGCTCCTGGGGTTCCTC
(13)
GGCAGCACAGTCGTCGCACATCTTGTCAACCTTCCATGCGAGGACCATCACGACTACCTTGGCTCCCTGGGGTTC
CTC
(14)
GGCAGCACAGTCGTCGCACATCTTGTCACCTTCCATGCGAGGACCATCTCGACGACCTGCTCCTGGGGTTCCTC
(15)
GGCAGCACAGTCGTCGCACATCTTGTCAACCTTCCATGCGAGGACCATCTCGACGACCTGCTCCTGGGGTTCCTC
(16)
GGCAGCCGGATCGTCGCACATCTTGTCAACCTTCCATGCGAGGACCATCTCGACGACCTGCTCCTGGGGTTCCTC
(17)
GGCAGCCGGATCGTCGCACATCTTGTCACCTTCCATGCGAGGACCATCTCGACGACCTGCTCCTGGGGTTCCTC
(18)
GGCAGCCGGATCGTCGCACATCTTGTCAACCTTCCATGCGAGGACCATCTCGACGACCTGCTCCTGGGGTTCCTC
(19) GGCAGCACAGTCGTCGCAATCCAGTCAAGACTGAACAGAGGACCATCTCGACGACCTGCTCCTGGGTT

(20)
GGCAGCACAGTCGTCGCACATCTTGTCAACCTTTTCCATGCGAGGACGAGCAGGACCTGGAACCTGGTGCGTTCC
TC
(21)
GGCAGCACAGTCGTCGCACATCTTGTCACCTTCCATGCGAGGACGAGCAGGACCTGGAACCTGGTGCGTTCCTC
(22)
GGCAGCACAGTCGTCGCACATCTTGTCAACCTTCCATGCGAGGACGATGCAGGACCTGGAACCTGGTGCGTTCCT
C
(23)
GGCAGCCGGATCGTCGCACATCTTGGTNAANCTTCCATGCGAGGACGAGCATGAACTGGAACCTGGTGCGTTCCT
C
(24) GGCGGATCGTCGCACATCTTGTCAACCTTCCATGCGAGGACCACGAGGTCTCCACTGGTCCAGGGGTTCCTC
(25)
GGCAGCACAGTCGTCGGCACATCTTTGGTCAACCTTCCATGCGAGGACCACGAGGTCTCCACTGGTCCAGGGTTC
CTC
(26)
GGCAGCCGGATCGTCGCACATCTTGTCAACCTTCCATGCGAGGACGACCAAGACGAGGACCAGGTAGGCGTTCCT
(27)
GGCAGCCGGAT423CGTCGCACATCTTGTCAACCTTCCATGCGAGGACGTGATGGAGCAAGTCCTCCTGTCGGTTC
CTC
(28)
GGCAGCACAGTCGTCGCACATCTTGTCAACCTTCCATGCGAGGACACGAGGTCTCCACTGGTCCAGGTTCCTC
(29)
GCCCAAACAAGTCGTCGCACATCTTGTCAACCTTCCATGCGAGGACCGAGNNNGTAGCTGGANNCTCGGATGCGT
TCCT
(30)
GCAGCACAGATCGTCGCACATGCTTGTCAAGCCTTTCCATCGCGAGGACCATCCTACGGAGCGAGCACTTGCTGC
CTGGGGTTC
(31)
GGCAGCCGGATCGTCGCACATCAATGGTTTGGCTGGTGATACTGAGGACCACGACGTCTACACTTGGTTCCAGGG
TTCCTC

These sequences could be translated into the following building block compositions:

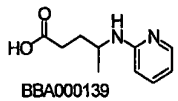
Sequence no.	Postion 1	Position 2	Position 3
1	BBA000098	BBA000085	BBA000100
2	BBA000098	BBA000099	BBA000100
3	BBA000152	BBA000099	BBA000100
4	BBA000153	BBA000152	BBA000100
5	BBA000098	BBA000099	BBA000139
6	BBA000098	BBA000099	BBA000139
7	BBA000098	BBA000099	BBA000139
8	BBA000098	BBA000099	BBA000139

9	BBA000098	BBA000099	BBA000139
10	BBA000098	BBA000099	BBA000139
11	BBA000098	BBA000099	BBA000139
12	BBA000098	BBA000099	BBA000139
13	BBA000098	BBA000099	BBA000139
14	BBA000098	BBA000099	BBA000139
15	BBA000098	BBA000099	BBA000139
16	BBA000152	BBA000099	BBA000139
17	BBA000152	BBA000099	BBA000139
18	BBA000152	BBA000099	BBA000139
19	BBA000098	BBA000088	BBA000139
20	BBA000098	BBA000099	BBA000053
21	BBA000098	BBA000099	BBA000053
22	BBA000098	BBA000099	BBA000053
23	BBA000152	BBA000099	BBA000053
24	BBA000152	BBA000099	BBA000140
25	BBA000098	BBA000099	BBA000140
26	BBA000152	BBA000099	BBA000167
27	BBA000152	BBA000099	BBA000098
28	BBA000098	BBA000099	BBA000200
29	BBA000098	BBA000099	-
30	BBA000098	BBA000099	-
31	BBA000152	BBA000160	-

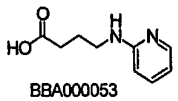
In position 1 L-Asp (BBA00098) dominated. D-Asp was also found (BBA000152)

In position 2 Gly (BBA00099) dominated.

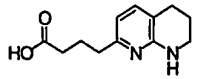
In position 3 building blocks carrying an amidine and no amine functionality was found to dominate:



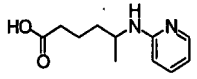
BBA000139



BBA000053

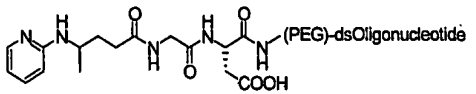


BBA000100



BBA000140

The most abundant sequence was thereby found to correspond to the following structure:



BBA000139-BBA000099-BBA000098

The following 3 sequences

BBA000098-BBA000099-BBA000139

BBA000098-BBA000099-BBA000100

BBA000098-BBA000099-BBA000053

out of the 31 identified sequences were selected for further analysis using an standard ELISA assay and thereby verified as binders of the αvβ3 Integrin receptor.

While the invention has been described with references to specific methods and embodiments, it will be appreciated that various modifications and changes may be made without departing from the invention. All patent and literature references cited herein are hereby incorporated by reference in their entirety.

Claims

1. A method for producing a composition of molecules with an improved desired property, comprising the steps of:

i) providing an initial library comprising a plurality of different encoded molecules associated with a corresponding identifier nucleic acid sequence, wherein each encoded molecule comprises a reaction product of multiple chemical entities and the identifier nucleic acid sequence comprises codons identifying said chemical entities,

ii) subjecting the initial library to a condition partitioning members having encoded molecules displaying a predetermined property from the remainder of the initial library,

iii) identifying codons of the identifier nucleic acid sequences of the partitioned members of the initial library, and

iv) preparing a second-generation library of encoded molecules using the chemical entities coded for by the codons of the partitioned members of the initial library or a part thereof.

2. The method according to claim 1, wherein the second-generation library comprises a plurality of different encoded molecules associated with a corresponding identifier nucleic acid sequence, wherein each encoded molecule comprises a reaction product of multiple chemical entities and the identifier nucleic acid sequence comprises codons identifying said chemical entities.

3. The method of claim 1 or 2, further comprising subjecting the second generation library to a condition partitioning members having encoded molecules displaying a predetermined property from the remainder of the second generation library.

4. The method according to any of the claims 1 to 3, further comprising the step of deducing the identity of the encoded molecule(s) using the identifier nucleic acid sequence.

5. The method according to claim 4, wherein the codons of the identifier nucleic acid sequence is decoded to establish the synthesis history of the encoded molecules.

6. The method according to any of the claims 1 to 5, wherein the encoded molecule associated with the corresponding identifier nucleic acid sequence is a bifunctional complex.

7. The method according to any of the claims 1 to 6, wherein the encoded molecule is covalently associated with the corresponding identifier nucleic acid sequence.

8. The method according to any of the claims 1 to 7, wherein the multiple chemical entities are precursors for a structural unit appearing in the encoded molecule.

9. The method according to any of the claims 1 to 8, wherein the chemical entities are reacted without enzymatic interaction to produce the encoded molecule.

10. The method according to any of the claims 1 to 9, wherein some or all chemical entities are not naturally occurring α-amino acids or precursors thereof.

11. The method according or any of the claims 1 to 10, wherein the encoded molecule is not an α-polypeptide.

12. The method according to any of the claims 1 to 11, wherein each codon comprises 4 or more nucleotides.

13. The method according to any of the claims 1 to 12, wherein the codons are separated by a framing sequence.

14. The method according to claim 13, wherein the framing sequence positions the reaction of a chemical entity in the synthesis history of the encoded molecule.

15. The method according to any of the claims 1 to 14, wherein the identifier nucleic acid sequence comprises two or more codons.

16. The method according to any of the claims 1 to 15, wherein the identifier nucleic acid sequence comprises three or more codons.

17. The method according to any of the claims 1 to 16, wherein the identifier nucleic acid sequence is amplifiable and comprises codons identifying chemical entities, which have participated in the formation of the encoded molecule.

18. The method according to any of the claims 1 to 17, wherein in step ii) the condition for partitioning of the desired library members includes subjecting the initial library to a molecular target and partitioning members binding to said target.

19. The method according to any of the claims 1 to 18, wherein the encoded molecule has a molecular weight less than 2000 Dalton, preferably less than 1000 Dalton, and more preferred less than 500 Dalton.

20. The method according to any of the claims 1 to 19, wherein the identifier nucleic acid sequence identifies the encoded molecule uniquely.

21. The method according to any of the claims 1 to 20, wherein the identifier nucleic acid sequence is detached from the encoded molecule.

22. The method according or any of the claims 1 to 21, wherein identifier nucleic acid sequence prior to step iii) is amplified.

23. The method of claim 22, wherein the identifier nucleic acid sequence is amplified applying the polymerase chain reaction (PCR).

24. The method according to any of the claims 1 to 23, wherein the codons of the identifier nucleic acid sequences of the partitioned members of the initial library are identified by contacting said identifier nucleic acid sequences with a pool of nucleic acid fragments under conditions allowing for hybridisation.

25. The method according to claim 24, wherein the pool of nucleic acid fragments comprises a plurality of single stranded nucleic acid probes immobilized in discrete areas of a solid support, wherein the nucleic acid probes are capable of hybridising to a codon of the identifier nucleic acid sequence comprising codons.

26. The method of claim 25, wherein the identity of the codons is revealed by observing the discrete areas of the support in which a hybridisation event has occurred.

27. The method according to any of the claims 24 to 26, wherein the nucleic acid probe of the array is hybridised to an identifier nucleic acid sequence through an adapter oligonucleotide having a sequence complementing the probe as well as one or more codons of the identifier nucleic acid sequence.

28. The method according to any of the claims 24 to 27, wherein a probe of the array is capable of hybridising to two codons of the identifier nucleic acid sequence or a sequence complementary to said sequence.

29. The method according to claims 24 to 28, wherein a nucleic acid probe of the array is capable of hybridising to all codons of an identifier nucleic acid sequence.

30. The method according to any of the claims 24 to 29, wherein a nucleic acid probe is capable of hybridising to all but one codon of the identifier, or less.

31. The method according to any of the preceding claims, wherein the existence of a hybridisation event is measured through labelling of the identifier nucleic acid sequence, or an amplification product thereof.

32. The method according to any of the claims 24 to 31, wherein the hybridisation event is measured by the emission of light in a scanner.

33. The method according to claim 31 or 32, wherein the relative intensity of light in each discrete spot is measured.

34. The method according to claim 24, wherein nucleic acid fragments are primer oligonucleotides, and the identification involves subjecting the hybridisation complex between the primer oligonucleotides and the identifier nucleic acid sequences to a

condition allowing for an extension reaction to occur when the primer is sufficient complementary to a part of the identifier nucleic acid sequence, and evaluating based on measurement of the extension reaction, the presence, absence, or relative abundance of one or more codons.

35. The method according to claim 34, wherein the condition, which allows for an extension reaction to occur, includes a polymerase or a ligase as well as suitable substrates.

36. The method according to claim 35, wherein the condition includes a polymerase and a substrate comprising a blend of (deoxy)ribonucleotide triphosphates.

37. The method according to any of the claims 34 to 36, wherein at least a part of the primer oligonucleotide is complementary to a codon.

38. The method according to claims 34 to 37, wherein at least a part of the primer oligonucleotide is complementary to a codon and an adjacent framing sequence.

39. The method according to any of the claims 34 to 38, wherein the sequence comprising the codon and an adjacent framing sequence has a total length of 11 nucleotides or more.

40. The method according to any of the claims 34 to 39, wherein the extension reaction is measured using the polymerase chain reaction (PCR), wherein the primer of claim 34 is involved in said PCR.

41. The method according to any of the claims 34 to 40, wherein a primer is labelled.

42. The method according to claim 41, wherein the primer is labelled with a small molecule, a radioactive component, or a fluorogenic molecule.

43. The method according to claim 42, wherein the small molecule label is selected from biotin, dinitrophenol, and digoxigenin, and the PCR amplicons are detected using an enzyme labelled streptavidin, anti-dinitrophenol. or anti-digoxigenin, respectively, reporter molecule.

44. The method according to any of the claims 34 to 43, wherein extension reaction is measured by real-time PCR.

45. The method according to claim 44, wherein the real-time PCR involves the use of an oligonucleotide probe responsible for the generation of a detectable signal during the propagation of the PCR reaction.

46. The method according to any of the claims 34 to 45, wherein the probe is designed to hybridise at a position downstream of a primer binding site.

47. The method according to claim 45 or 46, wherein the probe is a 5' nuclease oligoprobe or a hairpin oligoprobe.

48. The method according to claim 24, wherein the nucleic acid fragment is associated with a chemical entity precursor capable of being transferred to a recipient reactive group.

49. The method according to claim 48, wherein the pool of nucleic acid fragments comprise an anti-codon identifying the chemical entity, said anti-codon complementing a codon of one or more identifier nucleic acid sequences.

50. The method according to claim 48 or 49, wherein the pool of nucleic acid fragments further comprises anti-codons not complemented by codons on an identifier nucleic acid sequence.

51. The method according to any of the claims 48 to 50, wherein the nucleic acid fragments, each comprising an anti-codon and a chemical entity, hybridised to the identifier nucleic acid sequences comprising codons are recovered.

52. The method according or claim 51, further comprising formation of a second generation library of complexes, each member of the library comprising an encoded molecule and an identifier nucleic acid sequence, which codes therefore, using the recovered nucleic acid fragments as building blocks.

53. The method according to any of claims 48 to 52, wherein the identifier nucleic acid sequences of the complexes are recovered from the partitioned complexes of step ii) in claim 1.

54. The method according to claim 1, wherein the identifier nucleic acid sequences of the partitioned library members are amplified prior to the identification step.

55. The method according to claim 54, wherein the amplification is conducted using the polymerase chain reaction (PCR).

56. The method according to any of claims 1 to 55, wherein the identifier nucleic acid sequences comprising codons are immobilized during step iii).

57. The method according to 56, wherein, in step iii), the identifier nucleic acid sequences are immobilized on a solid phase and the pool of nucleic acid fragments is present in a mobile phase.

58. The method according to any of claims 1 to 57, wherein the conditions used during the contacting step allow for specific hybridisation between nucleic acid fragments and the identifier nucleic acid sequence comprising codons.

59. The method according to any of the claims 51 to 58, wherein the nucleic acid fragments are recovered using denaturing conditions.

60. The method according to any of the claim 1 to 59, wherein the second generation library is formed by

a) mixing under hybridisation conditions, nascent bifunctional complexes comprising a chemical entity or a reaction product of chemical entities, and an identifier nucleic acid sequence comprising codon(s) identifying said chemical entities, with the recovered nucleic acid fragments, said fragments comprising an oligonucleotide sufficient complementary to at least a part of the identifier nucleic acid sequence to allow for hybridisation, a transferable chemical entity and an anticodon identifying the chemical entity, to form hybridisation products,

b) transferring the chemical entities of the nucleic acid fragments to the nascent bifunctional complexes through a reaction involving a reactive group of the nascent bifunctional complex, in conjunction with a transfer of the genetic information of the anticodon.

61. The method according to claim 60, further comprising step c) separating the components of the hybridisation product and recovering the complexes.

62. The method of claim 60 or 61, wherein steps a) through c) are repeated as appropriate using the recovered complexes in step c) as the nascent bifunctional complexes in step a).

63. The method according to claims 60 to 62, wherein the genetic information of the anticodon is transferred by enzymatically extending the identifier nucleic acid sequence to obtain a codon attached to the bifunctional complex having received the chemical entity.

64. The method according to any of the claims 60 to 63, wherein the genetic information of the anticodon is transferred to the nascent complexes by hybridisation to a cognate codon of the identifier region.

65. The method according to any of the claims 60 to 64, wherein the second generation library are subjected to a partitioning according to step ii) of claim 1.

66. The method according to any of the claims 1 to 65, wherein, prior to the partitioning according to claim 65, the second generation library of complexes are contacted with sequences complementary to the identifier nucleic acid sequences, and the complexes which have hybridised with the complementary sequences are recovered and used in the method of claim 60.

67. The method according to claim 66, wherein the hybridisation product, prior to recovery of the complexes, is treated with an enzyme cleaving in the event a mismatch occurs.

68. The method according to claim 67, wherein the enzyme is selected from T4 endonuclease VII, T4 endonuclease I, CEL I, nuclease S1, or variants thereof.

69. The method according to any of the claims 1 to 68, wherein the second-generation library is prepared using chemical entities appearing in the initial library and chemical entities foreign to the initial library.

70. The method according to any of the claims 1 to 69, wherein the chemical entities used in the formation of the second-generation library occur in a concentration above a certain threshold in the partitioned library.

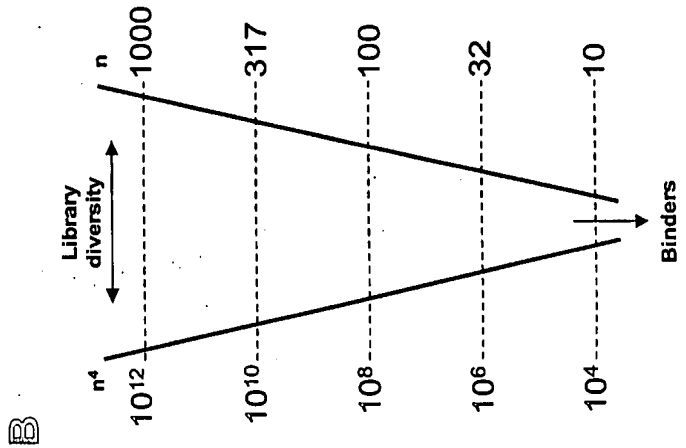
71. The method according to claim 70, wherein certain chemical entities occurring above a certain threshold is excluded in the second or further generation library.

72. A composition of molecules with an improved desired property, obtainable according to the method of any of the claims 1 to 71.

73. A molecule identifiable by subjecting a composition of molecules obtainable by a method according to any of the claims 1 to 71 to a condition partitioning members having encoded molecules displaying a predetermined property from the remainder of the composition, and identifying the partitioned encoded molecule(s).

74. The molecule according to claim 73, wherein the encoded molecule(s) are identified by decoding the identifier nucleic acid sequence.

75. The molecule according to claims 73 or 74, wherein the composition of molecules used in claim 73 is a second or further generation library.



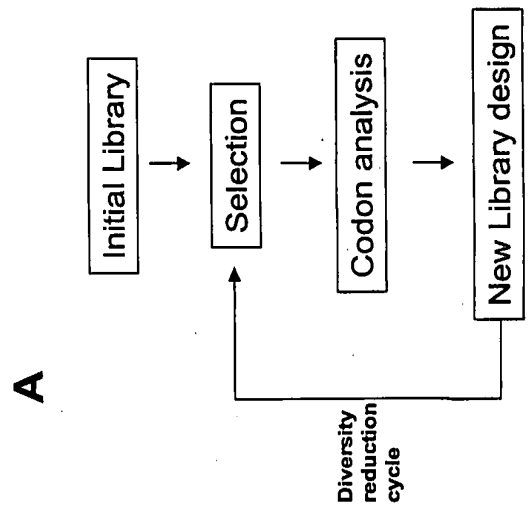


Fig. 1

Fig. 2

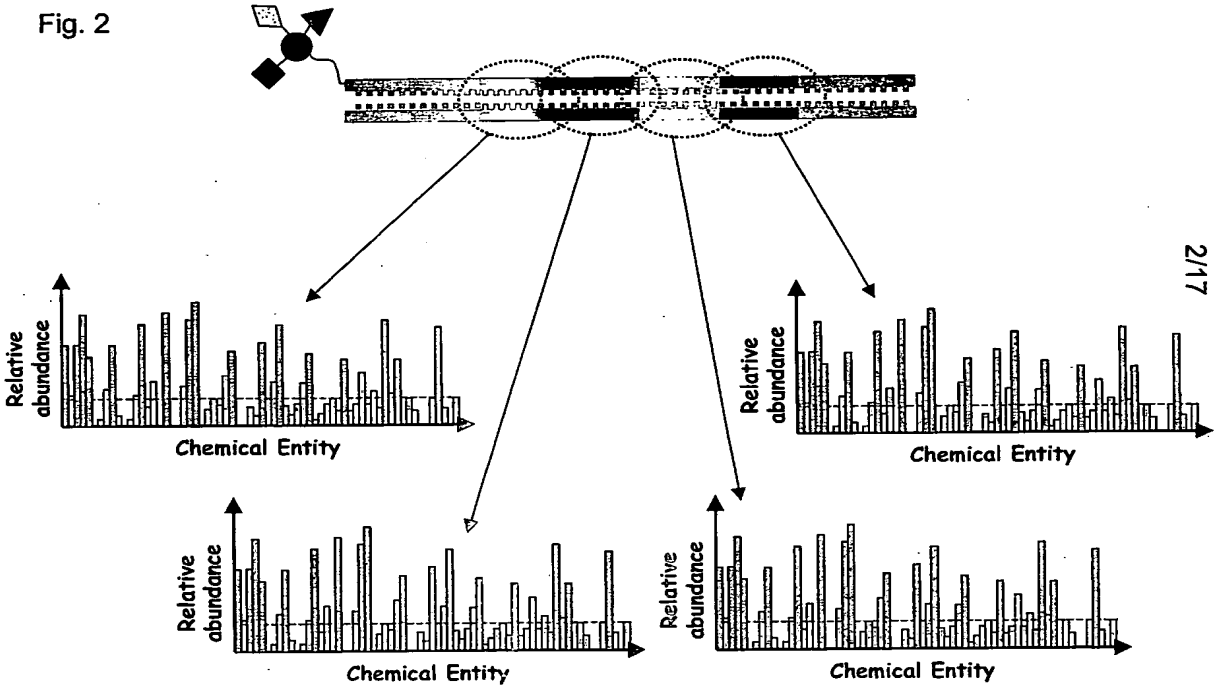

Relative abundance
Chemical Entity
Relative abundance
Chemical Entity
Relative abundance
Chemical Entity
Relative abundance
Chemical Entity

Fig. 3

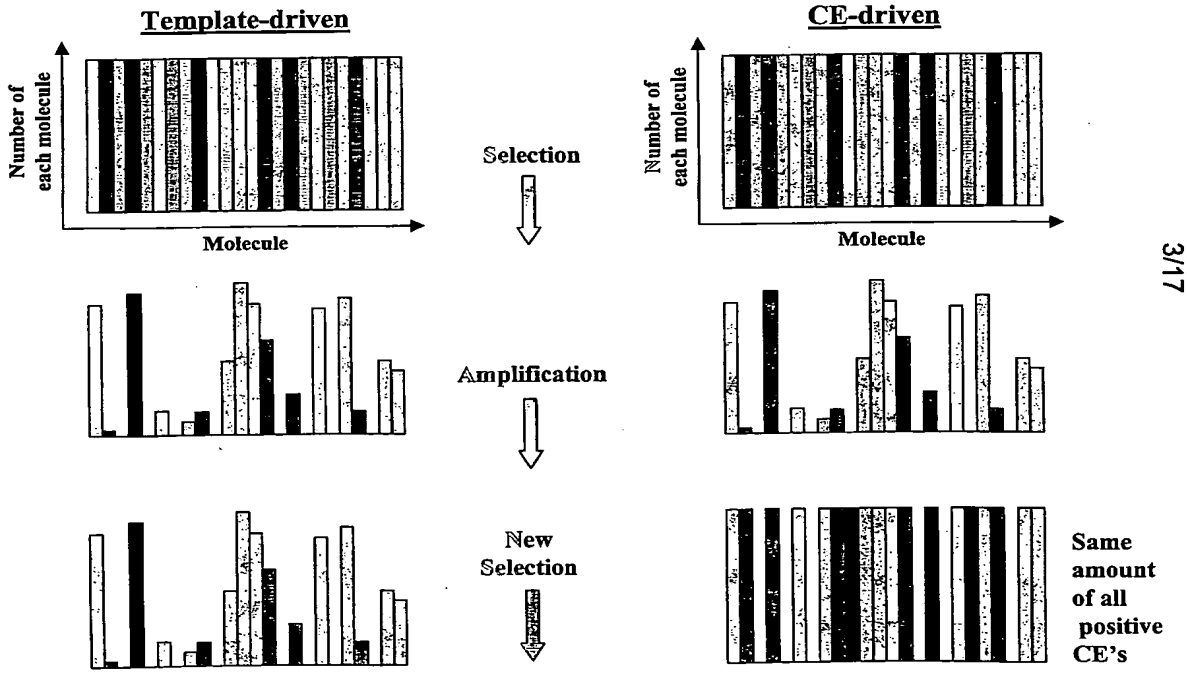

Template-driven
Number of each molecule
Molecule
Selection
CE-driven
Number of each molecule
Molecule
Amplification
New Selection
Same amount of all positive CE's

Fig. 4

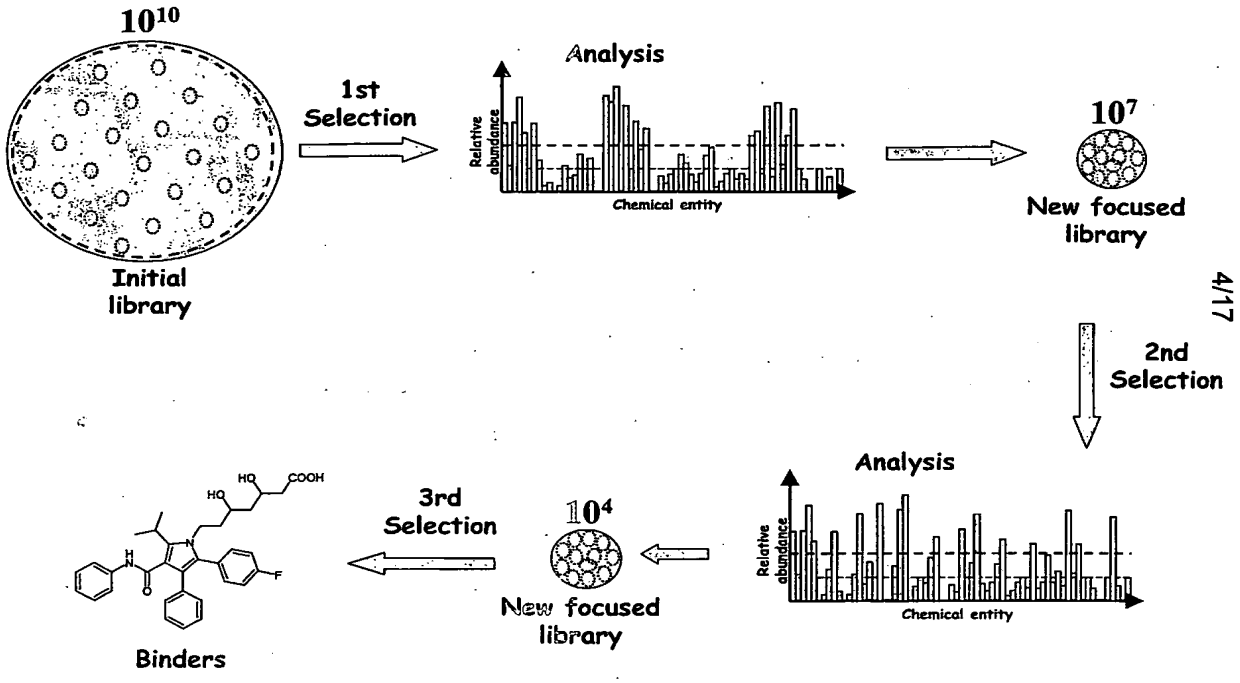
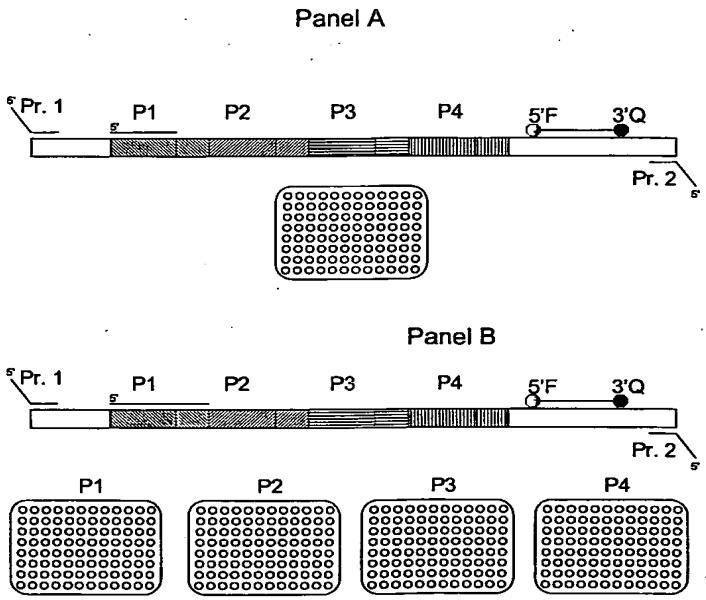


Fig. 5



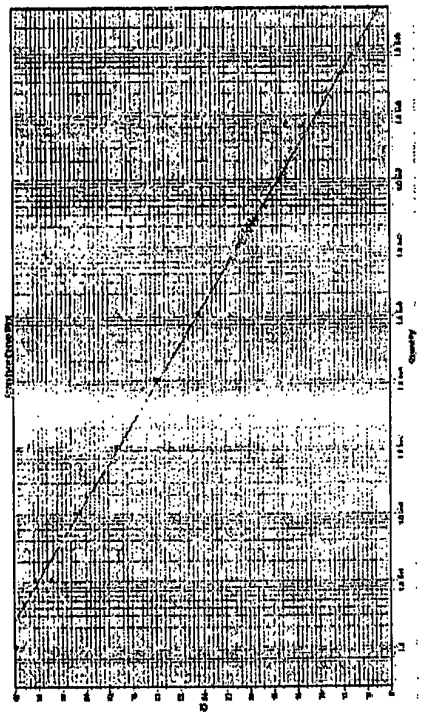


Fig. 6

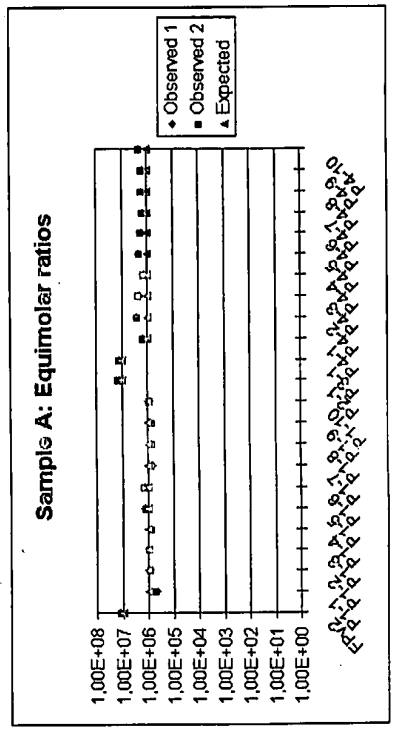

Sample A: Equimolar ratios
Observed 1
Observed 2
Expected
1,00E+08
1,00E+07
1,00E+06
1,00E+05
1,00E+04
1,00E+03
1,00E+02
1,00E+01
1,00E+00

Fig. 7

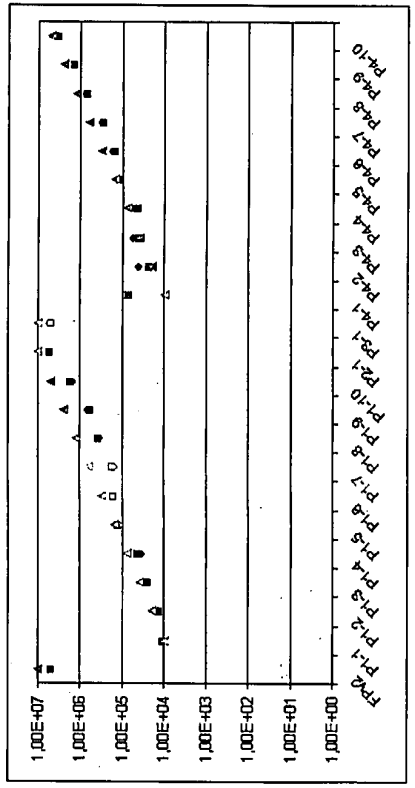

1,00E+07
1,00E+06
1,00E+05
1,00E+04
1,00E+03
1,00E+02
1,00E+01
1,00E+00
FPP2
PL-1
PL-2
PL-3
PL-4
PL-5
PL-6
PL-7
PL-8
PL-9
PL-10
PZL-1
PZ2-1
PK-1
PK-2
PK-3
PK-4
PK-5
PK-6
PK-7
PK-8
PK-9
PK-10

Fig. 8

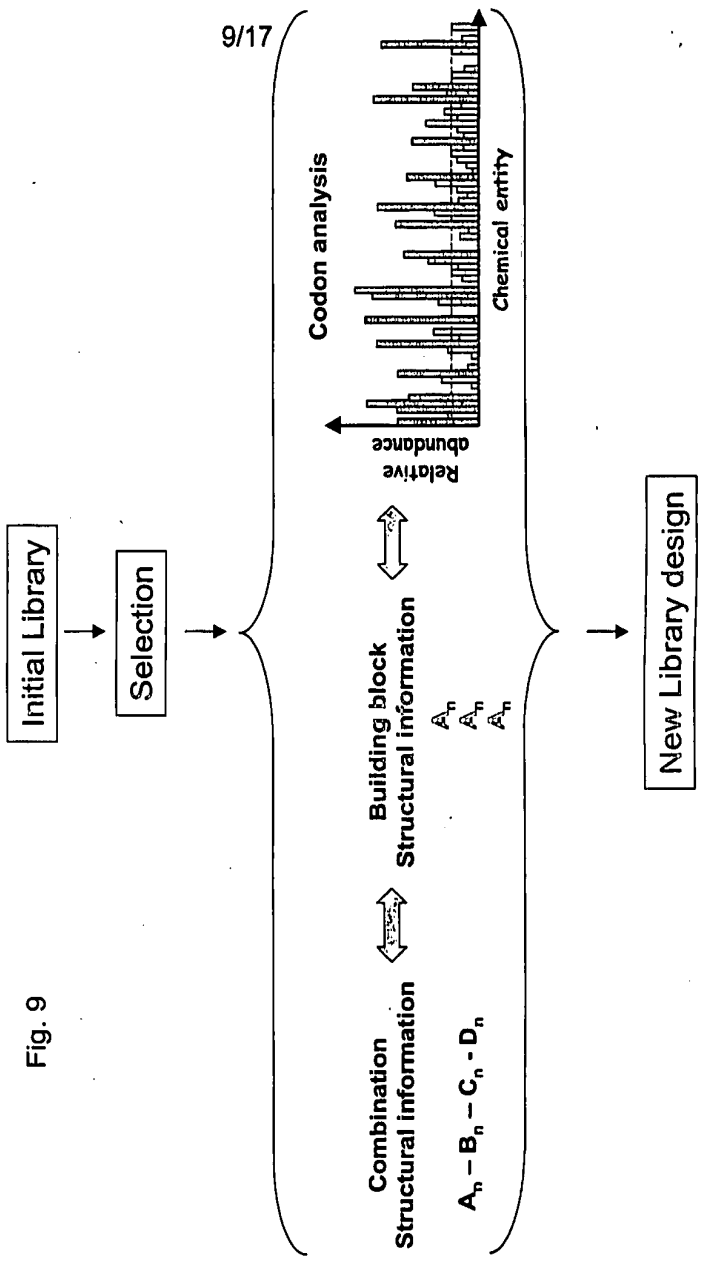

Initial Library
Selection
Codon analysis
Relative abundance
Chemical entity
Building block
Structural information
A_n
A_n
A_n
Combination
Structural information
$A_n - B_n - C_n - D_n$
New Library design

Fig. 9

Fig. 10

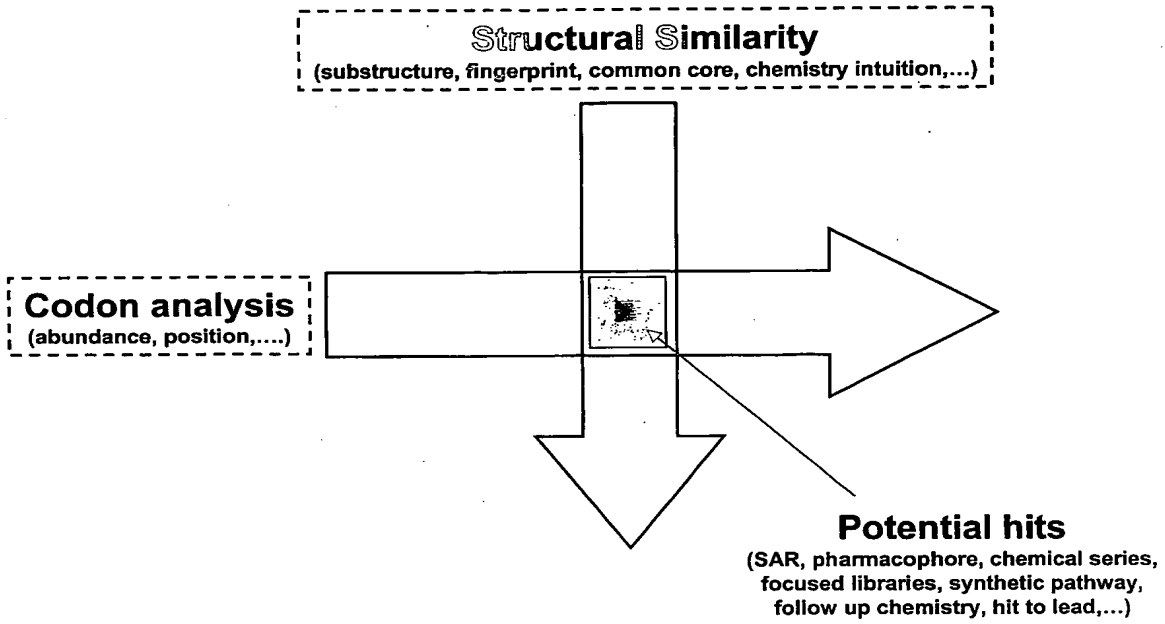
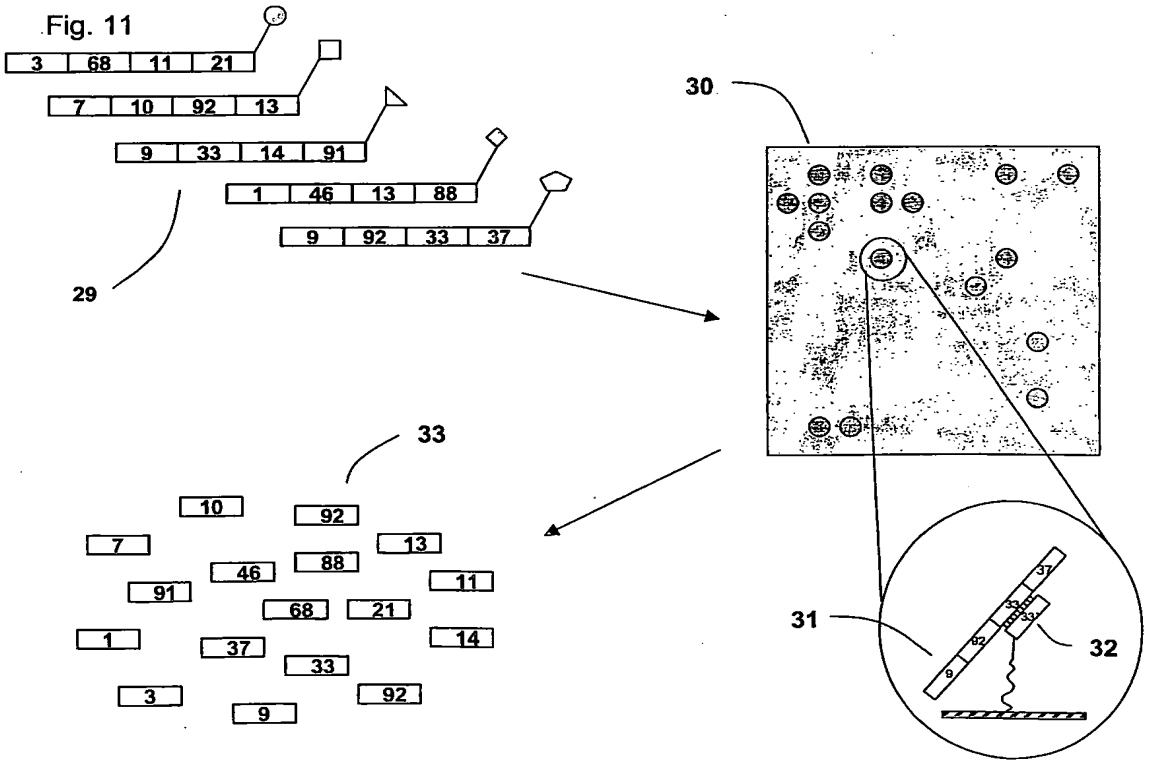

Structural Similarity
(substructure, fingerprint, common core, chemistry intuition,...)
Codon analysis
(abundance, position,....)
Potential hits
(SAR, pharmacophore, chemical series, focused libraries, synthetic pathway, follow up chemistry, hit to lead,...)

Fig. 11


3 68 11 21
7 10 92 13
9 33 14 91
1 46 13 88
9 92 33 37
29
30
33
10 92 7 13 46 88 91 11 68 21 1 14 37 33 3 92 9
31
32
37 33 33 92 9

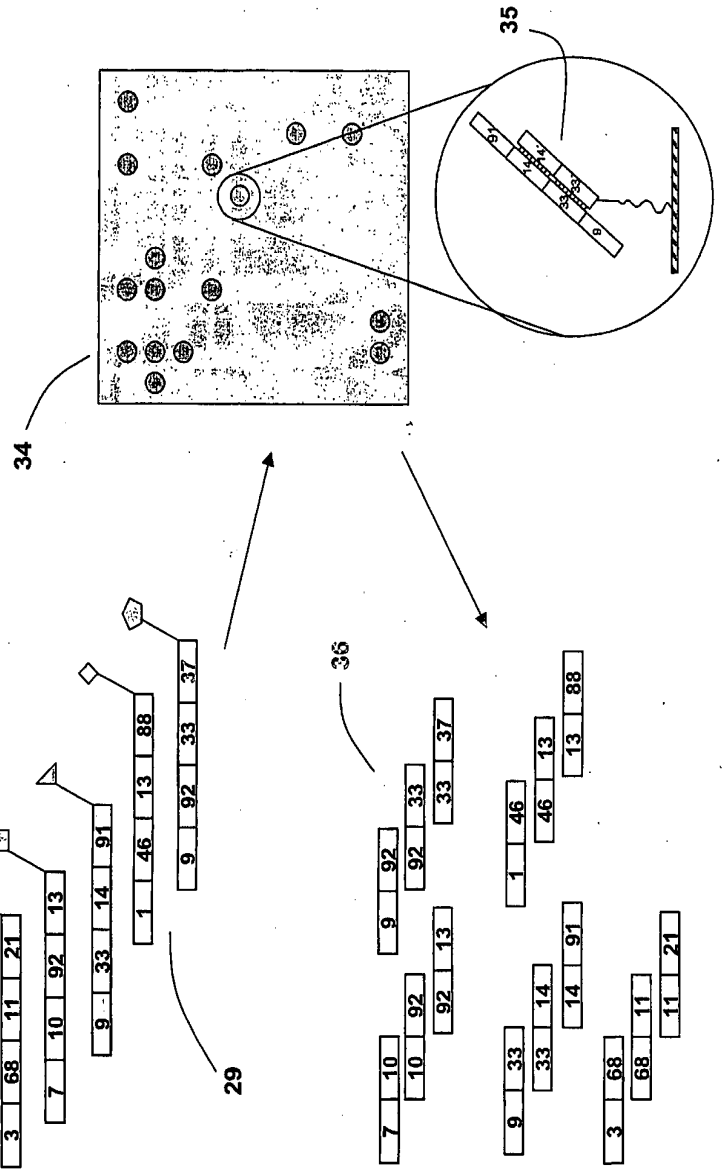


Fig. 12

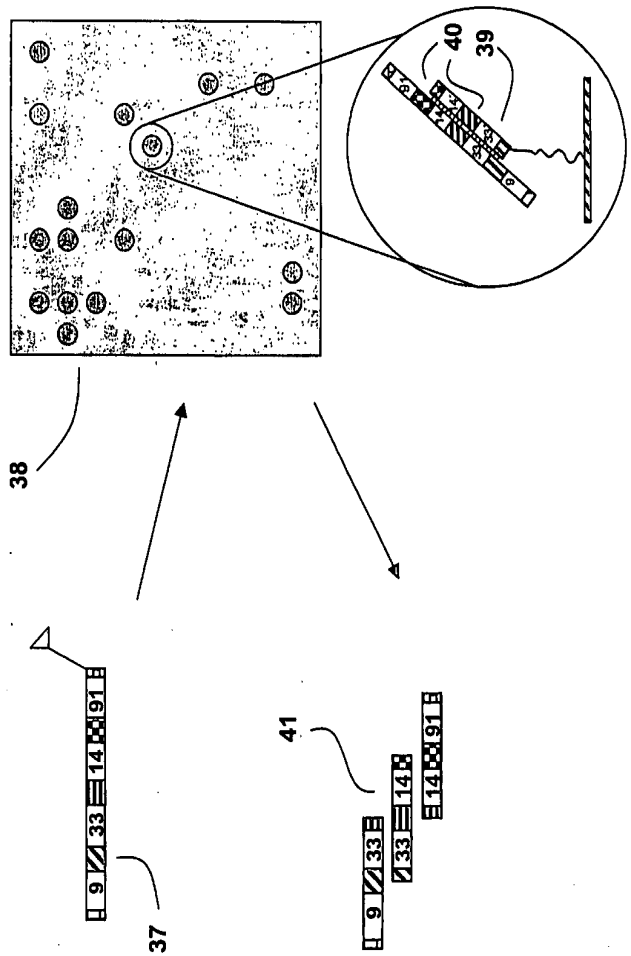


Fig. 13

Fig. 14

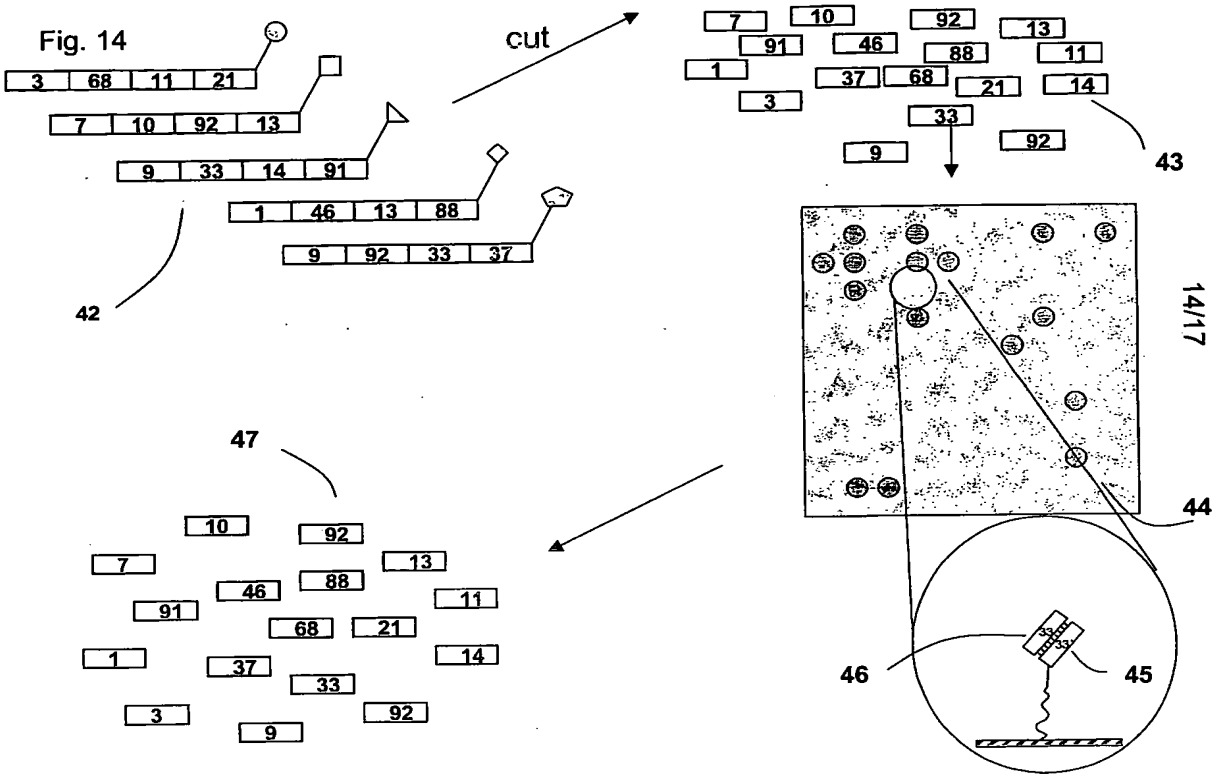
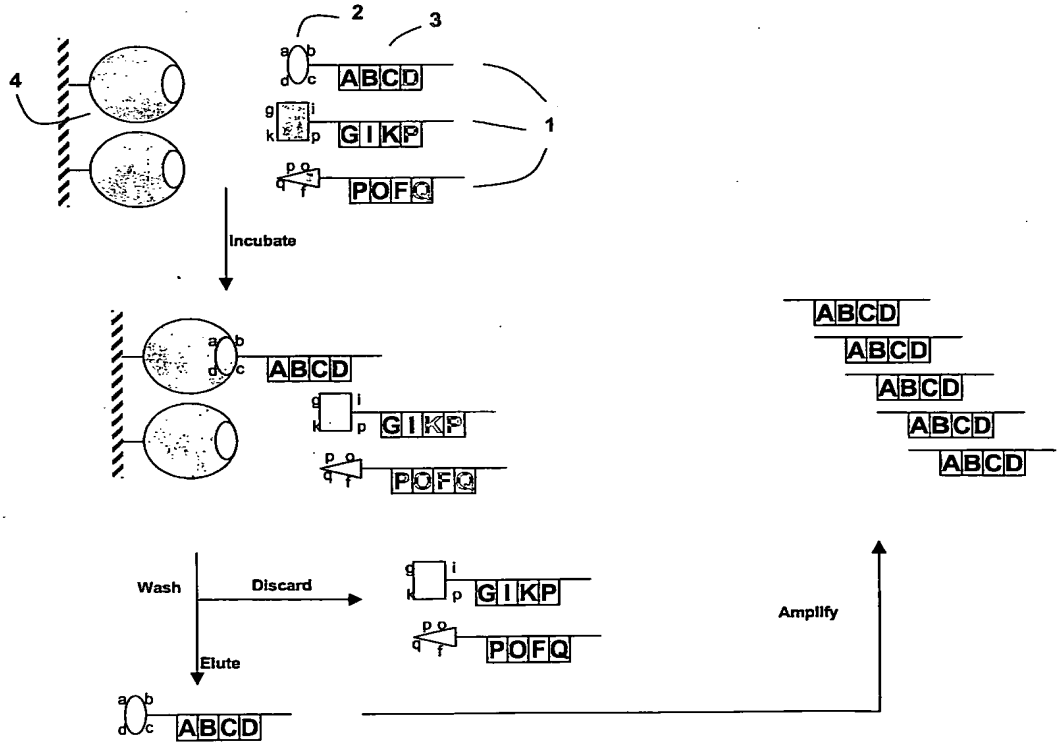


Fig. 15



WO 2004/074429 PCT/DK2004/000117

SUBSTITUTE SHEET (RULE 26)

Fig. 16

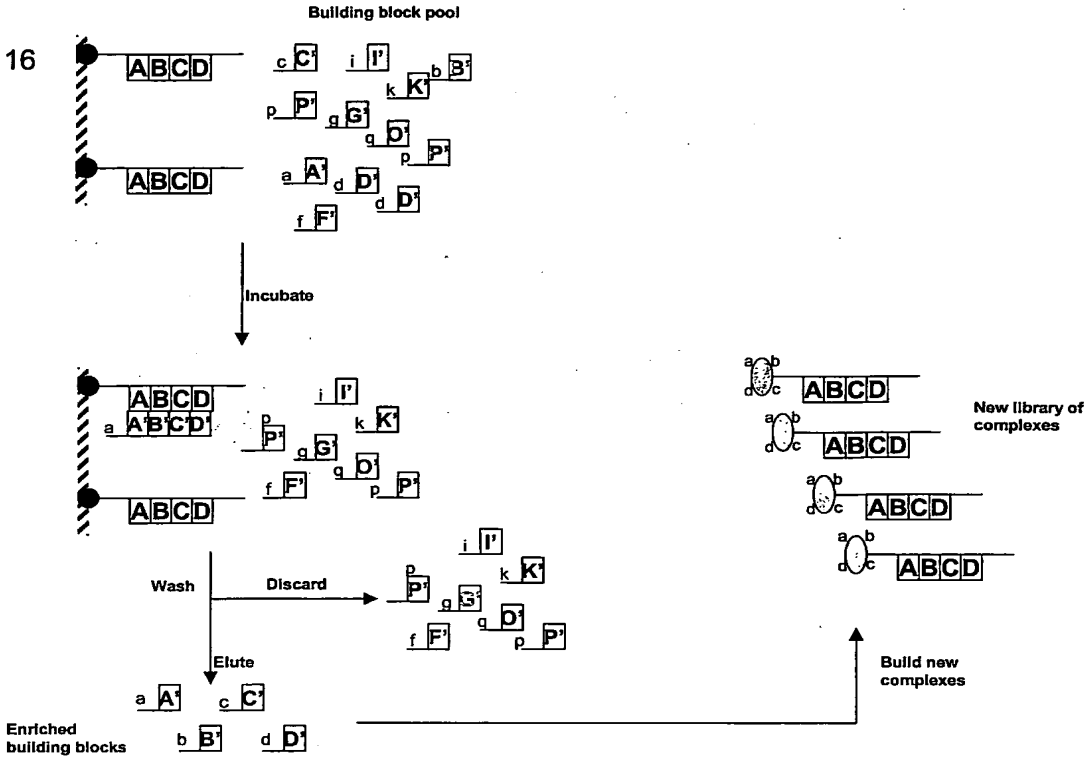
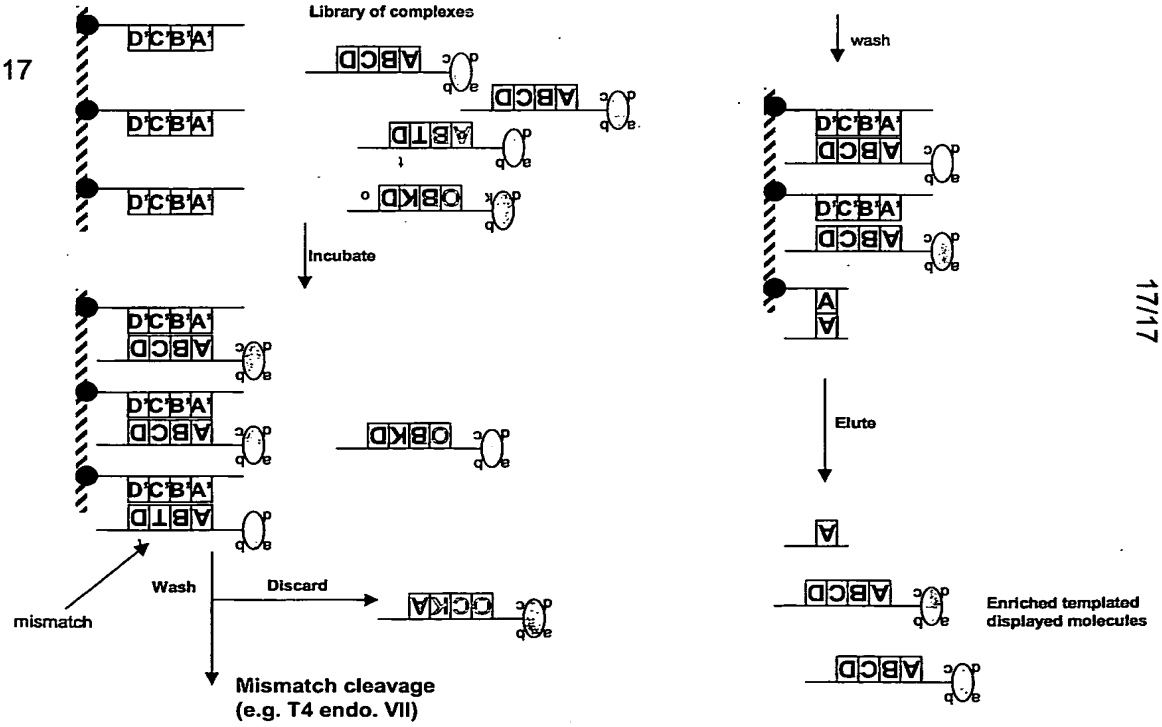


Fig. 17



(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
2 September 2004 (02.09.2004)

PCT

(10) International Publication Number
WO 2004/074429 A3

(51) International Patent Classification[7]: C12Q 1/68, C12N 15/10, C12P 1/00, C07B 61/00

(21) International Application Number:
PCT/DK2004/000117

(22) International Filing Date: 23 February 2004 (23.02.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

PA 2003 00268	21 February 2003 (21.02.2003)	DK
PA 2003 00269	21 February 2003 (21.02.2003)	DK
60/448,480	21 February 2003 (21.02.2003)	US
60/448,460	21 February 2003 (21.02.2003)	US
PA 2003 01356	18 September 2003 (18.09.2003)	DK
60/504,748	22 September 2003 (22.09.2003)	US

(71) Applicant (*for all designated States except US*): **NUEVOLUTION A/S** [DK/DK]; Rønnegade 8,5. sal, DK-2100 København Ø (DK).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **FRESKGÅRD, Per-Ola** [SE/SE]; Örtugsgatan 50, S-603 79 Norrköping (SE). **GOULIAEV, Alex, Haahr** [DK/DK]; Brødsted 223, DK-3670 Veksø Sjælland (DK). **THISTED, Thomas** [DK/DK]; Fjordskrænten 14, DK-3600 Frederikssund (DK). **OLSEN, Eva, Kampmann** [DK/DK]; Vingetoften 17, DK-2730 Herlev (DK).

(74) Agent: **HØIBERG A/S**; Store Kongensgade 59 A, DK-1264 Copenhagen K. (DK).

(81) Designated States (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declaration under Rule 4.17:
— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW, ARIPO patent (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)*

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

(88) Date of publication of the international search report:
30 September 2004

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: METHOD FOR PRODUCING SECOND-GENERATION LIBRARY

(57) Abstract: The present invention relates to a method for generating a second-generation library. In a first step, a library of encoded molecules associated with an identifier nucleic acid comprising codons identifying chemical entities that have participated in the formation of the encoded molecule is provided. In a second step, the library is partitioned and encoded molecules having a certain property are selected. Codons of identifiers of selected encoded molecules are subsequently identified, and a second-generation library is prepared using at least some of the chemical entities coded for by the identified codons. The new focussed library may be used for another partition step to select encoded molecules with a certain property.

WO 2004/074429 A3

INTERNATIONAL SEARCH REPORT

International Application No
PCT/DK2004/000117

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12Q1/68 C12N15/10 C12P1/00 C07B61/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12Q C12N C12P C07B

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, WPI Data, PAJ, MEDLINE, BIOSIS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 02/103008 A (GOUILAEV ALEX HAAHR ; NOERREGAARD-MADSEN MADS (DK); SLOEK FRANK ABILGA) 27 December 2002 (2002-12-27)	1-31, 34-43, 46, 48-66, 69-75
Y	the whole document	32,33, 44,45, 47,67,68
X	WO 02/074929 A (KANAN MATTEW W; GARTNER ZEV J ; LIU DAVID R (US); HARVARD COLLEGE (US)) 26 September 2002 (2002-09-26)	1-30, 34-43, 47-66, 69,72-75
Y	page 1 - page 81; claims 1-46; figures 1,3,5,11	31-33, 44-46, 67,68, 70,71

-/--

[X] Further documents are listed in the continuation of box C. [X] Patent family members are listed in annex.

* Special categories of cited documents :
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
13 August 2004	19/08/2004
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Authorized officer Botz, J

Form PCT/ISA/210 (second sheet) (January 2004)

INTERNATIONAL SEARCH REPORT

International Application No
PCT/DK2004/000117

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	GARTNER Z J ET AL: "Multistep small-molecule synthesis programmed by DNA templates" JOURNAL OF THE AMERICAN CHEMICAL SOCIETY, AMERICAN CHEMICAL SOCIETY, WASHINGTON, DC, US, vol. 124, no. 35, 4 September 2002 (2002-09-04), pages 10304-10306, XP002265219 ISSN: 0002-7863 the whole document	1-75
X	WO 00/23458 A (UNIV LELAND STANFORD JUNIOR) 27 April 2000 (2000-04-27) the whole document	1-75
Y	WO 93/20242 A (BRENNER SYDNEY ; SCRIPPS RESEARCH INST (US); JANDA KIM (US); LERNER RI) 14 October 1993 (1993-10-14) the whole document	1-75
Y	CALDERONE C T ET AL: "Directing otherwise incompatible reactions in a single solution by using DNA-templated organic synthesis" ANGEWANDTE CHEMIE, VCH VERLAGSGESELLSCHAFT, WEINHEIM, DE, vol. 41, no. 21, November 2002 (2002-11), pages 4104-4108, XP002282854 ISSN: 0044-8249 the whole document	1-75
Y	SUMMERER D ET AL: "DNA-templated synthesis: more versatile than expected" ANGEWANDTE CHEMIE, WILEY-VCH, WEINHEIM, DE, vol. 41, no. 1, 4 January 2002 (2002-01-04), pages 89-90, XP002265218 ISSN: 1433-7851 the whole document	1-75
P,X	WO 2004/013070 A (NUEVOLUTION AS ; PEDERSEN HENRIK (DK)) 12 February 2004 (2004-02-12) the whole document	1-75

Form PCT/ISA/210 (continuation of second sheet) (January 2004)

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/DK2004/000117

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
WO 02103008	A	27-12-2002	CA	2451524 A1	27-12-2002
			WO	02103008 A2	27-12-2002
			WO	02102820 A1	27-12-2002
			EP	1402024 A2	31-03-2004
			EP	1401850 A1	31-03-2004
			US	2003143561 A1	31-07-2003
			US	2004049008 A1	11-03-2004
			WO	03078625 A2	25-09-2003
			WO	03078445 A2	25-09-2003
			WO	03078626 A2	25-09-2003
			WO	03078050 A2	25-09-2003
			WO	03078446 A2	25-09-2003
			WO	03078627 A2	25-09-2003
WO 02074929	A	26-09-2002	US	2003113738 A1	19-06-2003
			CA	2441820 A1	26-09-2002
			EP	1423400 A2	02-06-2004
			WO	02074929 A2	26-09-2002
WO 0023458	A	27-04-2000	AU	1318400 A	08-05-2000
			CA	2346989 A1	27-04-2000
			EP	1123305 A1	16-08-2001
			WO	0023458 A1	27-04-2000
WO 9320242	A	14-10-1993	US	5573905 A	12-11-1996
			AT	193561 T	15-06-2000
			AU	685050 B2	15-01-1998
			AU	3944993 A	08-11-1993
			CA	2132103 A1	14-10-1993
			DE	69328781 D1	06-07-2000
			DE	69328781 T2	26-04-2001
			DK	643778 T3	07-08-2000
			EP	0643778 A1	22-03-1995
			ES	2147197 T3	01-09-2000
			JP	7505530 T	22-06-1995
			US	6060596 A	09-05-2000
			WO	9320242 A1	14-10-1993
			US	5723598 A	03-03-1998
WO 2004013070	A	12-02-2004	WO	2004013070 A2	12-02-2004

Form PCT/ISA/210 (patent family annex) (January 2004)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/DK2004/000117

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.b of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application and necessary to the claimed invention, the international search was carried out on the basis of:

   a. type of material

   - [ ] a sequence listing
   - [X] table(s) related to the sequence listing

   b. format of material

   - [ ] in written format
   - [ ] in computer readable form

   c. time of filing/furnishing

   - [X] contained in the international application as filed
   - [ ] filed together with the international application in computer readable form
   - [ ] furnished subsequently to this Authority for the purpose of search

2. [X] In addition, in the case that more than one version or copy of a sequence listing and/or table relating thereto has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that in the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments: